No. 87-666

FILED
JUL 13 1987
JOSEPH F. SPANIOL, JR.
CLERK

IN THE

Supreme Court of the United States

OCTOBER TERM, 1987

BRENDA BERKMAN, on behalf of herself and a class consisting of all similarly situated women,

Petitioner,

—v.—

THE CITY OF NEW YORK; EDWARD I. KOCH, individually and as Mayor of the City of New York; NEW YORK CITY FIRE DEPARTMENT; AUGUSTUS BEEKMAN, individually and as Fire Commissioner of the City of New York; NEW YORK CITY DEPARTMENT OF PERSONNEL; MICHAEL NADEL, individually and as Director of Personnel of the City of New York; THOMAS ROCHE, individually and as former Director of Personnel of the City of New York; CIVIL SERVICE COMMISSION OF THE CITY OF NEW YORK,

Respondents,

—and—

UNIFORMED FIREFIGHTERS ASSOCIATION and FIREFIGHTER ELIGIBLES ASSOCIATION, LIST NO. 1162, INC.,

Intervenor-Respondents.

APPENDIX TO PETITION FOR A WRIT OF CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE SECOND CIRCUIT

LAURA SAGER
Washington Square
Legal Services, Inc.
40 Washington Square South
Room 321B
New York, New York 10012
(212) 998-6176

Of Counsel

ROBERT L. KING
Debevoise & Plimpton
875 Third Avenue
New York, New York 10022
(212) 909-6000

Counsel of Record for Petitioner Brenda Berkman

July 13, 1987

EDITOR'S NOTE

THE FOLLOWING PAGES WERE POOR HARD COPY AT THE TIME OF FILMING. IF AND WHEN A BETTER COPY CAN BE OBTAINED, A NEW FICHE WILL BE ISSUED.

TABLE OF CONTENTS

	PAGE
Opinion of the United States Court of Appeals for the Second Circuit, February 17, 1987 (Berkman IV)	A1
Opinion of the United States District Court for the Eastern District of New York, October 8, 1985 (Berkman III)	A51
Order of the United States District Court for the Eastern District of New York, February 14, 1986	A102
Opinion of the United States Court of Appeals for the Second Circuit, March 29, 1983 (Berkman II)	A109
Opinion of the United States District Court for the Eastern District of New York, March 4, 1982 (Berkman I)	A169
Mandate of the United States Court of Appeals for the Second Circuit, February 17, 1987	A362

PAGE

Order of the United States Court of Appeals for the Second Circuit denying petition for rehearing, April 14, 1987 ..... A364

Opinion of the United States District Court for the Eastern District of New York, June 13, 1986 ..... A366

Civil Rights Act of 1964, as amended, 42 U.S.C. § 2000e (excerpts) A394

Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607 (excerpts) A397

UNITED STATES COURT OF APPEALS

FOR THE SECOND CIRCUIT

Nos. 1307, 1308, 1309,
1310--August Term, 1985

(Argued: May 28, 1986
Decided: February 17, 1987)

Docket Nos. 86-7157, -7159, -7167, -7201

BRENDA BERKMAN, on behalf of herself and a class consisting of all similarly-situated women,

Plaintiff-Appellee-
Cross-Appellant,

- v. -

THE CITY OF NEW YORK; EDWARD I. KOCH, individually and as Mayor of the City of New York; NEW YORK CITY FIRE DEPARTMENT, AUGUSTUS BEEKMAN, individually and as Fire Commissioner of the City of New York; NEW YORK CITY DEPARTMENT OF PERSONNEL; MICHAEL NADEL, individually and as Director of Personnel of the City of New York; THOMAS ROCHE, individually and as former Director of Personnel of the City of New York; CIVIL SERVICE COMMISSION OF THE CITY OF NEW YORK,

Defendants-Appellants-
Cross-Appellees,

-and-

UNIFORMED FIREFIGHTERS ASSOCIATION, LOCAL 94, FIREFIGHTERS ELIGIBLES ASSOCIATION, LIST NO. 1162, INC.,

Defendants-Intervenors-Appellants-Cross-Appellees,

-and-

JAMES T. AHRENS,

Defendant-Intervenor.

Before:

FEINBERG, Chief Judge, NEWMAN and KEARSE,

Circuit Judges.

Appeal and cross-appeal from orders of the District Court for the Eastern District of New York (Charles P. Sifton, Judge) providing supplemental relief in connection with a Title VII lawsuit alleging gender discrimination in entry-level hiring of New York City firefighters. 626 F. Supp. 591 (E.D.N.Y. 1985).

Affirmed in part, reversed in part, and remanded.

Norma Kerlin, New York, N.Y. (Frederick A.O. Schwarz, Jr., Corp. Counsel, Francis F. Caputo, Elizabeth Dale Kendrick, Robin M. Levine, New York, N.Y., on the brief), for municipal defendants-appellants-cross-appellees.

John F. Mills, Mineola, N.Y. (Colleran O'Hara & Mills, Mineola, N.Y., on the brief), for defendant-intervenor-appellant-cross-appellee Firefighter Eligibles Ass'n, List No. 1162, Inc.

Michael N. Block, New York, N.Y. (H. Adam Prussin, Cheryl Eisberg Moin, Lipsig, Sullivan & Liapakis, New York, N.Y., on the brief), for defendant-intervenor-appellant-cross-appellee Uniformed Firefighters Ass'n

Laura Sager, Washington Square Legal Services, Inc., New York, N.Y. (Robert L. King, Jonathan E. Richman, Debevoise & Plimpton, New York, N.Y., on the brief), for plaintiff-appellee-cross-appellant.

JON O. NEWMAN, Circuit Judge:

This is an appeal and cross-appeal from orders of the District Court for the Eastern District of New York (Charles P. Sifton, Judge) providing supplemental relief in connection with a Title VII lawsuit alleging gender discrimination in entry-level hiring of New York City firefighters. In an ear-

lier stage of this litigation, the District Court invalidated an entry-level examination and ordered various forms of relief, including the development of a new non-discriminatory entry-level test. The current round of litigation concerns challenges to the validity of the new test and to the District Court's orders requiring adjustments in the scoring of the new test and in the use of the eligibility list assembled as a result of the new test. For reasons that follow, we affirm in part, reverse in part, and remand for entry of a revised order.

Background

Much of the background is set forth in our prior decision, which rejected challenges to certain aspects of the relief the District Court had ordered after invalidating the physical portion of the original test. See Berkman v.

City of New York, 705 F.2d 584 (2d Cir. 1983). The plaintiff, Brenda Berkman, filed the suit in 1979, alleging gender discrimination by the New York City Fire Department and other municipal defendants in violation of Title VII of the Civil Rights Act of 1964, 42 U.S.C. § 2000e et seq. The plaintiff challenged the physical test of Exam 3040, the 1978 Fire Department entrance examination, on the ground that this test had a disparate impact on women and was not job-related. On March 4, 1982, the District Court invalidated the physical portion of Exam 3040 and ordered several forms of relief, including "preparation of new and valid selection procedures." Berkman v. City of New York, 536 F. Supp. 177, 216 (E.D.N.Y. 1982). This aspect of relief, which is customary in Title VII litigation, was not challenged on the prior

appeal. The March 4, 1982, decision also ordered as interim relief the hiring of up to 45 women members of the plaintiff class who passed a "qualifying test" of physical abilities. Such a test was developed by the defendants in cooperation with the plaintiff and approved by the District Court in August 1982. The qualifying test consisted of two parts, a simulation of engine company tasks and a simulation of ladder company tasks, separated by a rest interval. See 705 F.2d at 592 n.10. The test was scored on a pass/fail basis, with completion in four minutes, nine seconds, considered a passing score. This test was administered in September 1982. Thirty-eight of the women who passed were hired as firefighters.

On September 11, 1982, the defendants administered the written por-

tion of a new entry-level firefighter test, Exam 1162. The written portion was administered to 31,421 candidates of whom 566 identified themselves as females.

In October 1982 the defendants sought the District Court's approval of the physical test of Exam 1162. The physical test was similar to the "qualifying test" used for interim hiring, with some changes. Two additional tasks were added -- a hose pull and a wall vault. The rest interval between the engine company tasks and the ladder company tasks was reduced to two minutes. Finally, the scoring was altered from pass/fail to a rank-ordered system based on speed of completion. Completion in less than four minutes was scored 100, completion in each of the six 30-second intervals between four and seven minutes was scored downward from 95 to 70 in

five-point steps, and completion in more than seven minutes was considered failing. This produced seven passing grades or "bands." An applicant's overall score on Exam 1162 was to be determined by averaging the scores on the written and physical tests.

In January and February 1983 the District Court heard testimony on seven days concerning the validity of the physical test of Exam 1162. That hearing was adjourned on February 18, 1983, without a specific date for resumption. Two months later the defendants informed the Court that they were reluctantly going to administer Exam 1162, despite the lack of an advance ruling on its validity, because of the need to promulgate a new eligibility list and the unlikelihood that the hearing would be resumed and an advance ruling issued. Receiving no

contrary indication from the District Court, the defendants administered the physical test of Exam 1162, at a cost of $750,000, to more than 20,000 applicants who had passed the written test. Prior to administering the physical test, the defendants obtained foundation funding for a special training program for women to prepare them for the test. Most of the women who participated actively in the training program passed the physical test, with 40 percent scoring at least 85.

The physical test was administered during the period from July 1983 until the spring of 1984. During that period an episode occurred that would prove significant to one aspect of the remedy challenged on this appeal. In September 1983 the named plaintiff, Brenda Berkman, and another class member,

both of whom had been hired pursuant to the District Court's interim hiring remedy, were terminated at the conclusion of their probationary period, ostensibly for poor performance. This action precipitated a motion for reinstatement, plaintiff contending that the terminations had been the result of intentional discrimination. The District Court agreed and ordered reinstatement. Berkman v. City of New York, 580 F. Supp. 226 (E.D.N.Y. 1983).

The defendants disclosed the results of Exam 1162 in May 1984. Of the 31,421 applicants who took the written test, 29,113 achieved a passing score of at least 70. The passing rates were: men, 98.65 percent; women, 97.8 percent. The scores were bunched at the high end: 83.21 percent of the applicants scored 90 or higher, 66.68 percent scored 94 or

higher, and 9,788 applicants scored 98 or higher. Of the 28,559 men who passed the written test, 22,255 (77.93%) took the physical test; of the 554 women who passed the written test, 165 (29.78%) took the physical test. The passing rates on the physical test were: men, 95.42 percent; women, 46.67 percent. The distribution of scores on the physical test was as follows:

Score	Males	Females
100	500	0
95	6180	0
90	8529	7
85	3982	19
80	1325	18
75	451	25
70	169	8
below 70	1019	88

In June 1984 defendants disclosed a proposed eligibility list based on an equal weighting of the scores on the written and physical tests. The District Court estimated that approximately 2,800 applicants will be hired as

firefighters during the four-year life of the eligibility list, see N.Y. Civ. Serv. L. § 56 (McKinney 1983), and that approximately 6,500 applicants must be offered positions to obtain the needed 2,800. The 6,500 highest ranking applicants received a combined score of 94.5 or better. Only two women are in this group, which includes all applicants with any prospect of being hired as firefighters from this list.

In response to a motion by the plaintiff, the District Court issued an order on June 29, 1984, enjoining the use of the proposed eligibility list "until this Court has made a determination that Exam 1162 is valid and does not discriminate against women." An exception was made for interim hiring. The plaintiff reports that, as of March 1986, 850 fire-

fighters had been hired from the proposed list of eligibles, all of whom are male.

Between January and June 1985 the District Court conducted hearings on the validity of Exam 1162. The defendants sought to demonstrate the validity of the physical test of Exam 1162 on the basis of both content validity and criterion-related validity. Content validity concerns the measurement of knowledge or abilities needed for successful job performance. Criterion-related validity concerns the identification of criteria that reflect successful job performance and a determination of the extent to which test scores correlate with the meeting of such criteria. See Uniform Guidelines on Employee Selection Procedures (1978) of the Equal Employment Opportunity Commission, 29 C.F.R. § 1607.5(B), .14 (1986). The criterion-

related validation was based on a concurrent validation study, which compared test scores with job performances of a sample of 133 incumbent firefighters, 104 males and 29 females. Defendants' expert testified that his analysis showed a high degree of correlation for both males and females between physical test scores and job performance criteria.

The plaintiff's experts challenged the validity of Exam 1162 essentially on two grounds. First, they contended that the physical test measured a candidate's anaerobic energy system and ignored the aerobic system. Anaerobic energy is expended in using strength and speed for short intervals of time, usually less than five minutes. Aerobic energy is expended during physical exertion over prolonged periods of time. Weight-lifters and sprinters use primar-

ily anaerobic energy; long-distance runners use primarily aerobic energy. In the prior stage of this litigation, when the District Court invalidated Exam 3040, the physical portion of that exam had been criticized for testing only anaerobic energy, disregarding the fact that successful firefighting frequently requires paced exertion over several hours of activity. 536 F. Supp. at 207, 212. Plaintiff complained that Exam 1162 perpetuated this deficiency by requiring each of the two sets of physical tasks to be performed within 90 seconds in order to produce a score high enough to afford a candidate a realistic chance of being hired. Failure to test for stamina, it was urged, neglected a characteristic important for successful job performance and also slanted the scores adversely to women, who tend to compare more unfavor-

ably with men in regard to anaerobic energy than aerobic energy. Plaintiff's evidence indicated, for example, that men tend to run 40 percent farther than women in runs lasting two minutes, but only 14 percent farther in runs lasting ten minutes.

Second, plaintiff's experts challenged the use of scores from the written portion of Exam 1162. Noting that these scores were bunched at the high end, they contended that, because the written test was too easy, it did not provide sufficient differentiation among applicants. As a result, ranking on the proposed eligibility list was determined primarily by the results of the physical test, despite the fact that Fire Department officals had rated mental and physical abilities of equal importance for successful job performance.

In the course of presenting testimony challenging Exam 1162, one of plaintiff's experts suggested a way to rescore the results of the physical test by grouping the raw scores into three scoring bands, instead of the seven used by the defendants. The highest band included all who finished in less than four minutes 30 seconds, the second, those finishing between four minutes 31 seconds and six minutes; and the third, those finishing in more than six minutes one second but less than seven minutes. Rescoring of the physical test in three bands produced the following distribution:

Score	Males	Females
95 and 100	6,780	0
80, 85, and 90	13,836	44
70 and 75	620	33

For purposes of combining physical test scores with written test scores, the expert suggested assigning the top band a score of 100, the second band, 85, and the third band, 70. The expert urged that a three-band scoring would predict job performance as well as the seven-band scoring and would produce less adverse impact on women. Subsequently, the expert made it clear that his three-band proposal for the physical test scores would reduce adverse impact on women only if these scores were then combined with a computer-generated "normal distribution" of passing scores on the written test, randomly assigned to those who passed the written test.

On October 8, 1985, the District Court issued an opinion and order concerning the validity of Exam 1162. 626 F. Supp. 591 (E.D.N.Y. 1985). Judge

Sifton noted his prior criticism of Exam 3040 for its undue emphasis on anaerobic energy performance and concluded that in devising Exam 1162, "defendants failed lamentably to establish a basis for the emphasis placed on maximal strength and speed," id. at 598, and "ignored not only this Court's prior findings concerning the role of aerobic energy in performing firefighting functions but also the recommendations of the same expert whose Philadelphia evaluators appear to have been the principal source of both the qualifying exam and Exam 1162 that the test last 5 to 10 minutes without a recovery period," id. at 599 (footnote omitted). Nevertheless, the Court concluded that Exam 1162 "comports in general with the requirements of this Court's decision of March 1982 with the exception of its scoring which purports

to distinguish between qualified candidates to a degree of exactness not consistent with the lack of precision inherent in the exam." Id. at 593.

To remedy the scoring deficiencies, Judge Sifton directed three changes. First, he required that the physical test be scored in three bands. This system, he concluded, shows "greater validity" than the seven-band system and "appears required by defendants' own criterion measures as well as by the failure of defendants in their job analysis and test preparation to give due consideration to the demands made on aerobic energy in performing firefighting tasks," id. at 600 (footnote omitted). Second, Judge Sifton devised a remedy to deal with the fact that the percentage of those who passed the written test and went on to take the physical test was

much less for women than for men. The District Judge attributed this fall-off in interest to continued discrimination within the Fire Department as evidenced by the well-publicized efforts of the Department to discharge the plaintiff and another female probationary firefighter in September 1983. On the assumption that in the absence of the deterrent effect of the attempted firings, the percentage of those taking the physical test after passing the written test would have been the same for women as for men, Judge Sifton estimated that 432 women would have taken the physical test, instead of the 165 who did so, or 2.62 times as many. Id. at 600-01. On the further assumptions that, if 432 women had taken the physical test, they would have passed at the same rate as the 77 women who took and passed the test and

that all women who would have passed would have achieved a distribution of scores similar to those of the 77 who passed the test, Judge Sifton ordered that each woman on the eligibility list should be afforded an increased opportunity to be hired ahead of an equally ranked male. Id. at 601. The increased opportunity was to be achieved by use of a "compensation ratio" of 2.62 to 1. Third, to lessen the undue differentiating power of the physical test scores because of the bunching of the written test scores, Judge Sifton directed the parties to explore the validity and impact on women of rescoring the written test on either a pass/fail basis or in three scoring bands.

Each side presented entirely different proposals for a final order in response to the October 8 ruling. The

defendants presented evidence showing that use of a three-band scoring system for the physical test would have a more adverse effect on women than the seven-band system. Although only two women would be reached on the eligibility list for hiring with either scoring system, they would be reached later with use of the three-band system. Defendants therefore urged the District Court to permit use of the eligibility list as proposed. The plaintiff took the position that the Court's findings concerning the deficiencies in Exam 1162 required an extensive rescoring remedy. She proposed random selection from among all applicants who achieved a passing score on both the written and physical tests, with a female applicant accorded an increased opportunity to be hired over a male applicant in the ratio of 2.62 to 1. Alternatively,

she proposed that rank-ordering of candidates be permitted provided that male and female applicants were hired in the same proportion as would result from random selection of those who achieved passing scores. The plaintiff also offered a third and fourth alternative to be used in the event that the District Court was satisfied that the physical test had sufficient criterion-related validity to permit its use. The third alternative was to administer a new written test, presumably one with sufficient difficulty to produce a broader spread of test scores, and combine the scores on such a test with the scores from the physical test grouped in three bands. The fourth alternative was to generate by computer a "normal distribution" of passing scores for the written test, assign these scores randomly to all candidates who passed the

written and physical tests, and then combine these assigned scores with the scores from the physical test grouped in three bands.[1]

On February 14, 1986, the District Court issued a final order concerning Exam 1162. This order requires three changes in the scoring of Exam 1162 and the use of its results. First, the physical test is to be scored in three bands. Judge Sifton accepted the three bands as described in the testimony of one of

---

[1] A "normal distribution" is a pattern of the frequencies with which data occur at points along a continuum, the pattern resembling a bell-shaped curve and characterized by few occurrences at the low and high ends (in this case, scores near 70 and 100), the highest number of occurrences at the mean (in this case, 85), and nearly 70 percent of occurrences within one standard deviation of the mean (in this case, between 80 and 90). See R.P. Runyon & A. Haber, Fundamentals of Behavioral Statistics 112-13 (3d ed. 1977).

plaintiff's experts, placing scores of 100 and 95 in band A, scores of 90, 85, and 80 in band B, and scores of 75 and 70 in band C. However, the District Judge ordered that the scores to be combined with the written test scores would be 95.4 for candidates in band A, 87.6 for candidates in band B, and 73.6 for candidates in band C. Second, a "normal distribution" of test scores is to be computer generated for the written test and randomly assigned to all candidates who passed the written and physical tests. Third, the District Judge required use of the 2.62 compensation ratio outlined in the October opinion. The ratio is to be applied once a new rank-ordering of candidates has been compiled using the combined scores resulting from the random assignment of computer-generated written scores weighted equally with the three-

band scoring of the physical test. From such a list the defendants are to hire women and men as if there were 2.62 times as many women as actually appear at each level of the combined score.[2] The com-

---

[2] The Order details the use of the 2.62 compensation ratio as follows:

> Female applicants on the list shall be afforded an increased opportunity over equally ranked males on the list to be selected pursuant to a ratio of 2.62 to 1. This ratio shall be known as the compensation ratio. To apply this compensation ratio, the defendants shall divide the number of eligible men assigned a particular score on the list by the number of eligible women receiving the same score, times the compensation ratio of 2.62. The quotient of the calculation shall determine how many men receiving a particular score on the examination may be selected before the next woman is selected. This ratio shall be known as the selection ratio. Where men and women receive the same score, a woman shall be selected first, and the number of men determined by the selection ratio shall be selected thereafter. This process shall be repeated until the score is exhausted, and the next score shall be treated in the same manner, provided there [are] also women receiving that score.

pensation ratio gives each woman at any given score an increased chance of being selected over male candidates at that same score, though it does not require that any woman be hired ahead of any man with a higher score.

On the main appeal, the defendants challenge all three changes ordered by the District Court. On the cross-appeal, the plaintiff contends that the District Court should have required new written and physical tests, or at least should have required that the existing tests be used only for random selection from among all candidates who achieved passing scores.

Discussion

It will be useful to begin with consideration of plaintiff's cross-appeal, since if she is correct that the physical and written tests of Exam 1182

[sic] may not be used at all, there would be no need to consider the specific adjustments to the scoring of these tests ordered by the District Court. With respect to the written test, plaintiff has made no showing that would justify our rejection of this component of the exam. The written test was not claimed to have an adverse impact on women. Nor was it claimed to test for knowledge insufficiently related to job performance. Plaintiff's only complaint about the written test is that it is too easy. We will return to that claim in considering the defendants' objections to the scoring adjustments ordered by the District Court. At this point it suffices to note that the content of the written test has not been shown to be vulnerable on any ground cognizable under Title VII.

Plaintiff's challenge to the physical test is somewhat more substantial. She contends that the physical test has a demonstrably adverse impact on women and that it has not been adequately validated to justify its use. Her primary complaint is that the test measures an applicant's anaerobic energy system, as applied to firefighters' tasks, and substantially ignores the aerobic energy system. Assessment of the latter, she alleges, would afford female applicants an improved opportunity to achieve higher scores compared to those of male applicants. In considering this argument, we note first that, though the District Court expressed some criticism of the physical test for inadequate assessment of anaerobic [sic; aerobic] energy capacity, the Court nonetheless upheld the validity of the test. Having reviewed

the record and the decision of the District Court made on the basis of that record, we are not persuaded that the decision upholding validity was erroneous. The test was carefully constructed after a detailed consideration of firefighters' tasks. Substantial testimony before the District Judge supported validity on the basis of both content validation and criterion-related validation. Though plaintiff finds fault with the methods whereby the defendants demonstrated both content and criterion-related validity, there is an insufficient basis to disregard the District Court's conclusion upholding the validation of the test.

We do not doubt the plaintiff's basic point that stamina, a function of a person's aerobic energy system, is important in the performance of a fire-

fighter's tasks. The evidence of senior officials of the Fire Department acknowledged that stamina was an important attribute for successful job performance. It does not follow, however, that a physical test of the ability to perform simulated job tasks of firefighters, without a specific measurement of stamina, lacks validity to a degree that renders it vulnerable to a Title VII challenge. Obviously, firefighters frequently face situations where their anaerobic abilities determine whether or not they will save the lives of fire victims. The firefighter arriving on the scene of a fire will frequently be obliged to use strength and speed in a short amount of time. Abundant evidence in the record supports this point, which in any event would be self-evident. It may well be that the effectiveness of a person with

minimal stamina will decline if called upon to perform firefighting tasks over a considerable period of time. Perhaps a person with greater stamina would perform the tasks better after protracted activity than the firefighter who might excel in the first few minutes of activity. But the Fire Department is entitled to select those who are endowed with the physical abilities to act effectively in the first moments of arrival at a fire scene, where immediate speed and strength literally concern matters of life and death. If a person with limited stamina tires during the course of firefighting duties, that person can be replaced with a fresh firefighter. However, if the first firefighters on the scene are deficient in the speed and strength necessary to handle their tasks, those in need of immediate rescue will not be comforted by

the fact that those first on the scene might be able to sustain their modest energy levels for a prolonged period of time. See Spurlock v. United Airlines, Inc., 475 F.2d 216, 219 (10th Cir. 1972) (employer's burden to justify employment criteria correspondingly lighter where "human risks involved").

In an ideal world, a fire department might first select those applicants with a high degree of speed and strength and from that group make a second selection of those with relatively greater stamina. There is nothing in this record, however, to show that such a selection process would have a less adverse effect upon women. Indeed, since only seven women placed in the top 15,316 applicants on the physical test, which primarily measured speed and strength in the performance of firefighters' tasks, a

further selection from among these applicants, giving priority to those with relatively greater stamina, would at most have placed only these seven women applicants somewhat higher on the eligibility list, an outcome by no means certain.

In sum, the District Court's conclusion that the written and physical tests of Exam 1162 are appropriate for use to select entry-level firefighters is entitled to be approved.

We turn then to the defendants' challenges to the three changes in scoring ordered by the District Court. The first adjustment -- replacement of the seven-band system for scoring the physical test with a three-band system -- is fatally flawed. In the first place, the defendants have demonstrated, without contradiction by the plaintiff, that the three-band scoring system does not ad-

vance any objective of Title VII: it neither enhances the validity of the physical test nor reduces the adverse effect upon women applicants. In fact, it operates to the detriment of both test validity and women applicants. As we pointed out above, speed in the performance of a firefighter's task is highly relevant to successful job performance; collapsing the seven bands of test completion times into three bands serves only to oblige the defendants to select some applicants ahead of others who have demonstrated the capacity to handle firefighter tasks more swiftly. From the standpoint of women applicants, the three-band system does not place even one additional woman applicant high enough on the eligibility list to have any prospect of being reached for selection and in fact postpones the time when the two

women high enough to be selected will be appointed. Plainly, this remedy is unwarranted.

The second adjustment -- rescoring the written test -- arises from a trilemma faced by the District Court arising from two undisputed facts. The Fire Department officials evaluated both cognitive and physical skills as important for successful job performance. In addition, scores on the written test were bunched at the high end, according the physical test more differentiating power in the ultimate selection of candidates than the written test. Since the defendants were entitled to use the distribution of scores on the physical test in the selection of applicants, there were essentially three possibilities for use of the written test. The first was the plaintiff's preference for a new written

test of greater difficulty, which would produce a broader distribution of scores reflecting the range of cognitive abilities of those taking a more difficult test. The second was rescoring the written test to eliminate bunching of scores at the high end; of various techniques available, the District Court chose generating by computer a normal distribution of passing test scores and assigning such scores randomly to all who passed the written and physical tests. The third was the defendants' preference to leave the written scores unadjusted.

Each approach has some deficiency. Use of a more difficult written exam would encounter the substantial argument that cognitive abilities have been differentiated to a degree greater than that required for successful job performance. See Vulcan Society v. Civil

Service Commission, 360 F. Supp. 1265, 1276 (S.D.N.Y.), aff'd in part, remanded in part, 490 F.2d 387 (2d Cir. 1973). Though cognitive ability is important for successful performance as a firefighter, it does not follow that extremely high degrees of cognitive ability that might be measured by a difficult written test will provide a basis for selecting more competent firefighters. Use of the District Court's remedy encounters a different problem. Those who scored significantly better on the written test than other applicants and thereby demonstrated somewhat better cognitive ability within a range of abilities appropriate for hiring selection will be deprived of the competitive advantage they earned. Moreover, use of a normal distribution of scores on the written test may give this test more differentiating power than the

physical test, since scores on the latter did not follow a normal distribution pattern. Use of the defendants' solution also is not without a deficiency. Leaving the written scores unadjusted accords enhanced differentiating power to the scores of the physical test.

Though the facts created a trilemma, they did not warrant the District Court's remedy of random assignment of written test scores based on a computer-generated normal distribution of scores. That remedy unfairly deprives many male and female applicants of the enhanced opportunity they achieved by scoring comparatively better on the written test than other applicants. Moreover, it burdens the Fire Department with the prospect of hiring some applicants who, though achieving passing scores on the written test, ranked below other

applicants and thereby demonstrated a lesser degree of cognitive ability. Though the one-point differences at the high end of the score distribution probably lack significance, see Guardians Ass'n v. Civil Service Commission, 630 F.2d 79, 100-05 (2d Cir. 1980), cert. denied, 452 U.S. 940 (1981), the distribution of written scores used by the defendants spans scores throughout the range from 70 to 100. Even though scores were concentrated at the high end, whatever differentiating power the written test has could not be eliminated unless substantially justified to avoid noncompliance with Title VII. Such was not the case. The deficiency the District Court sought to avoid was according greater differentiating power to the physical test than to the written test. Though that outcome may have placed male

applicants higher on the eligibility list than they would have been had a normal distribution of written test scores been used, this consequence did not impair the validity of the physical test nor that of Exam 1162 as a whole. As discussed above, the defendants have an entirely legitimate interest -- a job-related interest -- in according priority in hiring to those with the demonstrated ability to perform firefighting tasks speedily. Indeed, the defendants would have been entitled, had they chosen, to score the written test solely on a pass/fail basis, using its results only as a threshold to identify the group from which applicants would then be selected on the basis of job-related physical abilities. The defendants did not violate Title VII by letting the results of the written test exert some differentiat-

ing power on the final eligibility list though less than that of the physical test.

The third scoring adjustment -- use of a 2.62 compensation ratio to enhance the hiring opportunity of women at the same combined score on the eligibility list as men -- is also unwarranted. Affirmative relief that accords enhanced hiring opportunities to compensate for the effects of past discrimination is available only under limited circumstances. See Local 28, Sheet Metal Workers' Int'l Ass'n v. E.E.O.C., 106 S. Ct. 3019, 3050 (1986) (plurality opinion); Kirkland v. New York State Department of Correctional Services, 711 F.2d 1117, 1134 (2d Cir. 1983), cert. denied, 465 U.S. 1005 (1984). In this case, Judge Sifton adopted an affirmative remedy because he concluded that publicity sur-

rounding the Fire Department's attempt to discharge the plaintiff and another female probationary firefighter had deterred women who had passed the written test from returning to take the physical test. This conclusion was an inference drawn from the facts that the terminations of the two probationary firefighters occurred shortly before the September 12-13, 1983, dates during which women applicants took the physical test, the terminations received considerable publicity, and the percentage of those passing the written test who showed up to take the physical test was much lower for women than for men.

In assessing whether this inference of deterrence may support a compensating hiring ratio, we note first the numerical context in which the ratio has been imposed. Defendants' projections

indicate that, even with all of the scoring adjustments ordered by the District Court, only six women will be ranked high enough on the eligibility list to be offered appointments during the life of the list. Since we have concluded that the three-band scoring of the physical test (which aided no women) and the random assignment of a normal distribution of scores on the written test are not warranted, the number of women who will likely be offered appointment remains at two, the original estimate when the list was first assembled. Since the compensation ratio applies only to increase the hiring priority of women at the same score on the list as men, rather than to give women a higher score, its effect is extremely limited. It will only enable the two women likely to be offered appointment to join the Fire Department

sooner. Thus, this is the rare case where an affirmative remedy has an extremely minimal effect. That circumstance would appear to reduce both the objection to the remedy and the need to adopt it.

Several considerations persuade us that the use of the affirmative remedy of a compensation ratio is not warranted in this case. There is no testimony from any female who passed the written test that she was deterred from taking the physical test by discriminatory conduct of the Fire Department. In this respect, the circumstances are quite unlike those that we previously found sufficient to support interim preferential hiring after Exam 3040 was invalidated. Several women testified that the difficulty of that test, later invalidated, had deterred them from taking it. Moreover, prior to

any publicity concerning the two terminations, the defendants provided a special training program to assist female applicants in scoring well on the physical test. The approximately 300 women who chose not to participate in this program could not have been deterred by any publicity concerning the two terminations, since the training program began two months before the publicity. In addition, the fact that the defendants made the training program available and secured foundation funding to support it militates against a conclusion that an affirmative remedy is warranted to overcome a demonstrated hostility to female candidates. Finally, it is significant that the defendants have offered to afford female candidates who passed the written test an additional opportunity to take the physical test and have further

offered to provide a training program to enhance their chances of success. If, as Judge Sifton concluded, some women were deterred from taking the physical test because of the two terminations, the defendants' offer is a more promising remedy than the 2.62 compensation ratio. Providing training for the approximately 300 women who did not take the physical test and giving them a second chance to take the test offers at least the prospect that some may score high enough to secure a combined score that will enable them to join the only two women now high enough on the list to be appointed. This is a more useful remedy than simply advancing the time of appointment for the two women now likely to be hired.

Conclusion

For all of these reasons, we affirm the October 8, 1985, and the February 14, 1986, orders of the District Court to the extent that they uphold the validity of Exam 1162; we reverse the orders to the extent that they require three-band scoring of the physical test, random assignment of computer-generated scores for the written test, and use of a compensation ratio; the defendants may promptly issue and make appointments from an eligibility list compiled from the combined scores on the written and physical tests, without adjustments; that list shall be supplemented with the combined scores of any women who have passed the written test and accept the defendants' offer to participate in a training program and take the physical test again.

The orders of the District Court are affirmed in part, reversed in part, and remanded for entry of a revised order consistent with this opinion.

Brenda BERKMAN, Plaintiff,

v.

The CITY OF NEW YORK et alia, Defendants.

No. CV-79-1813

United States District Court
E.D. New York

October 8, 1985

Women's Rights Clinic, Washington Square Legal Services, by Laura Sager, and Debevoise & Plimpton, by Robert L. King, Bart Schwartz, New York City, for plaintiff.

Booth, Lipton & Lipton, by Charles J. Hynes, New York City, for plaintiff-intervenor Ahrens.

Frederick A.O. Schwarz, Corp. Counsel of the City of N.Y., by Norma Kerlin, Elizabeth Dale Kendrick, New York City, for City defendants.

Colleran, O'Hara & Mills, P.C., by John F. Mills, Mineola, and Lipsig, Sullivan & Liapakis, by Michael N. Block, New York City, for defendant-intervenor UFA Local 94.

Salles & Danzig, by Stuart Salles, New York City, for defendant-intervenor Firefighters Eligibles Ass'n List #1162, Inc.

AMENDED MEMORANDUM
AND ORDER

SIFTON, District Judge.

This action was commenced in 1979 to remedy alleged discrimination against women in the selection of New York City Firefighters in violation of Title VII of the Civil Rights Act of 1964, as amended, 42 U.S.C. § 2000e et seq.; the Civil Rights Act of 1871, as amended, 42 U.S.C. § 1983; the fourteenth amendment to the United States Constitution; and the New York Human Rights Law.

Plaintiff sued on her own behalf and as representative of a class of women who took and passed the written portion of the qualifying entry-level test for firefighters administered in 1977 ("Exam 3040") and, thereafter, either took and failed or were deterred from taking the physical portion of that test. On July 10, 1980, this Court de-

ermined that the action could be main-
ained by plaintiff on behalf of a class
f women so described.

Trial of the action occurred in
981, and on March 4, 1982, this Court
etermined that Exam 3040 violated Title
II. Berkman v. City of New York, 536
. Supp. 177 (E.D.N.Y. 1982), aff'd, 705
.2d 584 (2d Cir. 1983). As a result,
he Court directed defendants to hire up
o 45 class members found to be qualified
ursuant to a special qualifying examina-
ion to be agreed upon between the par-
ies or, in the absence of agreement, to
e determined by the Court. In addition,
he Court directed defendants to commence
orthwith preparation of a new qualifying
xam that did not discriminate against
omen. On March 25, 1982, the Court
ntered a supplemental order directing
he parties to cooperate in an effort to

each an accord as to the new selection procedure.

In August of 1982 the court approved an agreement between the parties with respect to the qualifying test to be used as a basis for selecting the class members to be appointed to the fire department pursuant to the Court's prior decision. This test was administered to interested class members in September 1982, and from the class members passing the examination, some 38 women became firefighters.

Efforts to reach agreement with regard to the content of a new physical test for all entry-level firefighters were less successful. Accordingly, in October 1982, defendants presented a proposed new test to the Court for approval, and in January and February 1983 evidence was presented with respect to

the validity of the new test. At the conclusion of seven days of testimony on the subject, the Court adjourned proceedings without a date to permit the parties to explore further the possibilities of arriving at an agreement as to the contents of the new test. By May 1983, with no agreement reached, defendants, out of concern for the age of the list of persons eligible for appointment created from the previous Exam 3040,[1] proposed to proceed with the administration of the new test (Exam 1162), which thereafter occurred over a period of several months.

During this same period of time, in September 1983, plaintiff and

---

[1] Pending the creation of a new exam, appointments of male firefighters continued to be made from the eligibility list created from Exam 3040, with corresponding increases in the number of positions open to females who passed the qualifying exam.

nother class member, Zaida Gonzalez, ere terminated at the conclusion of heir probationary period ostensibly ecause of their poor performance as robationary firefighters. A motion made y plaintiff to compel both class mem- ers' reinstatement was heard by this ourt in October 1983, which, on December , 1983, found that the two women had een subjected to intentional discrimina- ion on account of their sex and directed heir reinstatement. Berkman v. City of ew York, 580 F. Supp. 226 (E.D.N.Y. 983).

Following the completion of the dministration of the new test, plaintiff oved in June 1984 to enjoin the promul- ation of a new eligibility list derived rom it pending the completion of hear- ngs with regard to the validity of the xam. On July 6, 1984, this Court en-

tered an order restraining the promulga-
tion of the eligibles list and appoint-
ments from it without prior court order,
upon a showing of compelling necessity,
pending determination of the validity of
the exam. After further efforts to re-
solve the issues presented by the new
test were aborted, hearings resumed and
continued through the spring of 1985.

Based on the evidence adduced
at the hearings with respect to the new
exam, I conclude that it comports in
general with the requirements of this
Court's decision of March 1982 with the
exception of its scoring which purports
to distinguish between qualified candi-
dates to a degree of exactness not con-
sistent with the lack of precision inher-
ent in the exam. Accordingly, defendants
are authorized to promulgate an eligibles
list created from Exam 1162 and to make

appointments therefrom in the manner set forth below. What follows sets forth the findings of fact and conclusions of law on which these determinations are based as required by Rule 52(a) of the Federal Rules of Civil Procedure.

1. Standing. The first issue requiring consideration is the standing of plaintiff to seek a determination that Exam 1162 does not comply with this Court's orders directing the preparation and use of a new physical exam for entry level firefighters which does not discriminate against women. I conclude plaintiff has standing.

The decree which plaintiff seeks to enforce is one she obtained in protracted litigation in which she has consistently taken an active role both to secure her own rights and those of other class members. No suggestion has been

ade that she lacked standing to obtain he portions of the decree she now seeks o enforce, see Allen v. Wright, [468 .S. 737,] 104 S. Ct. 3315 (1974); ilmore v. City of Montgomery, 417 U.S. 56, 570 n.10 (1974). Nor has any argu- ent been advanced that this Court lacks he jurisdiction retained in the original ecree "to secure compliance with [the ourt's original] order . . . [and] for onsideration of such further interim or emedial relief as may be necessary and ppropriate." See, e.g., Green v. County chool Board, 391 U.S. 430, 439 (1968); uardians Ass'n of New York City Police, nc. v. Civil Service Commission of New ork, 630 F.2d 79, 109 (2d Cir. 1980), ert. denied, 452 U.S. 940 (1981); Davis . Board of Education, 674 F.2d 684, 680 8th Cir.), cert. denied, 459 U.S. 881 1982).

Both objectively and subjec-
ively, plaintiff has demonstrated a real
nd concrete interest in assuring that
he and her fellow class members have a
on-discriminatory work environment in
hich to carry out their functions as New
ork City firefighters, an environment
hich would be substantially undermined
f the next class of women applicants for
he position of entry-level firefighters
re deprived of an opportunity to join
he ranks because of their sex.

As already noted, there are now
ess than 40 women firefighters in New
ork City out of a total of approximately
en thousand. Because of the nature of
he department's organization, no more
han one women firefighter has been as-
igned to any firehouse. As also previ-
usly noted, plaintiff herself and at
east one other female firefighter were,

after their appointment, subjected to intentional discrimination based on their sex in the all-male firehouses to which they were assigned. See Berkman v. City of New York, supra, 580 F. Supp. 226. Thus, plaintiff and the class she represents face the prospect of losing much if not all they gained if, as a result of discrimination in further hiring, they are left as an insular minority of women firefighters who "slipped through the [Department's] allegedly discriminatory screening practices." See Gray v. Greyhound Lines East, 545 F.2d 169, 173, 175-76 (D.C. Cir. 1976); Chicano Police Officers' Ass'n v. Stover, 526 F.2d 431, 437 (10th Cir. 1975), vacated on other grounds, 426 U.S. 944 (1976), standing aff'd, 522 F.2d 918, 921 (10th Cir. 1977).

Contrary to the argument of defendants and intervenors, this is not a case like Allen v. Wright, supra, 104 S. Ct. at 3331, in which a plaintiff asserts no more than the stigmatic injury which she shares with any other woman seeking to work for the New York City fire department. Here, plaintiff has suffered injury not only in being denied appointment by virtue of a prior discriminatory entrance exam, but thereafter by discrimination in the essentially male-only workplace in which she functions, resulting in her termination. By her efforts to ensure that Exam 1162 is non-discriminatory, she seeks to ensure that she will prserve the gains she has won as a result of the earlier trial. The presence of increased numbers of qualified women in the workplace will demonstrate, in a fashion in which no

udicial decree can be expected to, that ostility towards women firefighters is isplaced and furnish witnesses ready and illing to speak out against continued iscrimination by defendants and their mployees.

2. Disparate Impact. This ourt's earlier decree required defen- ants to prepare a new physical exam for ntry-level firefighters which did not iscriminate against women. Accordingly, t is appropriate to consider whether the esults of the exam present a prima facie ase of sex discrimination and, if so, hether that prima facie appearance of iscrimination may be explained by the equirements of the job. See Griggs v. uke Power Co., 401 U.S. 424, 432 (1971); lbemarle Paper Co. v. Moody, 422 U.S. 05, 425 (1975); Guardians Ass'n, supra, 30 F.2d at 88.

Here, the undisputed facts concerning the hiring that may be expected to result from the administration of Exam 1162 demonstrate a disparate impact on women directly attributable to the physical portion of the facially neutral test instrument because of the statistically disproportionate results for men and women.

The written portion of Exam 1162 was administered on September 11, 1982, to 31,421 candidates of whom 28,946 identified themselves as males and 566 identified themselves as females. 30,474 candidates or 98.58 percent passed the test, representing 98.65 percent of the male applicants (28,559) and 97.8 percent of the female applicants (554). In this connection, it bears noting that, despite a determination that the cognitive abilities tested by the written portion of the

xam should be weighted equally with the esults of the physical exam in determining which applicants were most qualified o be firefighters, the written exam ailed lamentably to distinguish between he varying cognitive abilities of the pplicants passing the exam. 83.21 percent of the candidates received a score f 90 or higher; 66.68 percent received a core of 94 or higher; the median score as between 96 and 97. 9,788 candidates eceived a score of 98 or above.

The physical portion of Exam 162 was administered over several months o those candidates passing the written est with all female applicants being ested on two days, September 12 and 13, 983. Unfortunately, the testing of the omen applicants coincided with extensive ublicity concerning discrimination on he job against plaintiff and her fellow

lass member Gonzalez, which eventually esulted in what this Court found, after earings, to be retaliatory discharges on eptember 21, 1983 of the two women.

As occurred with respect to the rior physical exam, 3040, a surprisingly mall proportion of the women passing the ritten exam presented themselves to take he physical exam. Thus, while 22,255 en, or 77.93 percent of those eligible, ook the physical exam, only 165 women, r 29.78 percent of the eligible women, id so. Had women candidates eligible to ake the exam presented themselves in the ame proportion as did their male counerparts, 432 women, rather than 165, ould have taken the test.

Out of the total of 21,313 andidates, both men and women, taking he physical exam, 21,236 men and 77 omen passed it. The pass rate for men

was 95.42 percent; the pass rate for women was 46.67 percent. Thus, the pass rate for women on the physical exam was less than 80 percent, i.e., less than four-fifths of the pass rate for men, and the pass rates for the two groups were separated by at least 27 standard deviations. Accordingly, it is clear that the physical exam had a disparate impact on women. See Castenada v. Partida, 430 U.S. 482, 497 n.17 (1977); EEOC Uniform Guidelines on Employee Selection Procedures (the "Guidelines"), 29 C.F.R. § 1607.4(D).

3. Relation to the Job. As was the case in the trial of the underlying action, the parties have not attempted to prove all that the job of firefighting in New York City consists of so as to compare it with the test to establish whether the test is, in fact,

job related. Instead, the parties have concentrated their efforts on the issue of test preparation in order to argue whether the methods of test preparation engaged in by the parties were or were not of the sort from which a job-related test could be reasonably expected to result. The distinction is significant because in the case of Exam 1162, unlike Exam 3040, all parties to a greater or lesser extent participated in the test preparation and did so with substantial amount of job knowledge and sophistication about job testing, acquired as a result of months of involvement in this litigation. Plaintiff's efforts to measure the validity of Exam 1162 solely by reference to defendants' compliance with a variety of standards for the preparation of neutral tests, including the EEOC Guidelines, the American Psychological

Association's Standards for Educational and Psychological Tests (1974) (the "APA Standards") and its Principles for the Validation and Use of Personnel Selection Procedures (2d ed. 1980) (the so-called "Division 14 Principles") ignores the very substantial role she and her experts and attorneys played in the preparation of Exam 1162. And, while plaintiff points to the drawbacks of preparing Exam 1162 in the context of ongoing litigation in terms of the antagonism she and women applicants for employment faced as a result of her role in bringing to light sex discrimination in the fire department, she fails to place proper emphasis on the opportunities she enjoyed as a result of the adversarial process to prevent discriminatory impulses from realizing themselves in any blatant way in the process of test preparation.

As plaintiff concedes the qualifying test pursuant to which she and a sizeable portion of the class interested in becoming firefighters in fact did so "established the framework for Exam 1162." Given the remarkable degree of success with which the class members performed on the qualifying test and the opportunities afforded plaintiff to improve upon it in developing a new entrance exam, it is difficult to accept plaintiff's position that so much of Exam 1162 as derives from the qualifying test is not job-related and valid.

The qualifying test proposed by the City to this Court on May 25, 1982, consisted of two parts, a simulation of engine company tasks and a simulation of ladder company tasks, separated by a 5-minute rest interval. The engine company simulation consisted of two tasks: drag-

ging a length of 3-1/2 inch hose and lifting a roll of 2-1/2 inch hose from the ground and carrying it upstairs. The ladder company simulation consisted of five tasks: raising a portable ladder, climbing a preset portable ladder with a hook, entering a window and climbing stairs to the fifth floor, hitting an 80 pound rolled hose with an 8 pound maul down the length of a 6 foot table, and dragging a 145 pound dummy around a marked course. The City proposed a cutoff time of 4 minutes, 9 seconds to complete the test.

There can be little doubt that the qualifying test was developed under the pressures of litigation without great attention to the guidelines, standards and principles available to test preparers proceeding on a more leisurely basis, and, yet, there is little reason to think

that the test proposed did not, in the main, adequately test for the important behavioral requirements of the job.

For one thing, the qualifying test, unlike Exam 3040, did not purport to test for the abstract physical endowments necessary to do firefighting work. Instead, it was founded on tasks in many important respects like those engaged in by firefighters on the job.[2] Whatever the merits of a test of this sort in situations in which only one or none of

---

[2] In fact, the principal source of the City's proposed test appears to have been a criterion test used by the City of Philadelphia to measure performance by incumbent firefighters. While there are, of course, dangers in using a job performance measure to choose between applicants for entry-level employment, those dangers appear to have been in the main eliminated here because of the simplicity of the test as modified and the familiarization program afforded applicants prior to the test's administration.

the parties has previously studied the important work behaviors representative of the job, in a situation such as that presented here in which all involved had been studying that subject exhaustively as a result of ongoing litigation, a job-sample test had the advantages of both facial validity (securing acceptance of test results both from those who take it and those for whose benefit it is taken) and recognized content validity for those who have become job experts in the course of ongoing litigation.

In fact, the City's proposed qualifying examination was speedily content-validated in this sense shortly after its initial presentation. Plaintiff's expert, Dr. Raymond Mendel, after a week of revisiting City firehouses and observing job behavior with particular attention to the tasks identified as

important in the City's proposed test, criticized several aspects of the City's proposal including the forcible-entry simulation, the rolled-hose carry, and the pace at which the test was to be performed. As a result of these criticisms, the parties agreed to modifications of the test by having the candidates lift the 2-1/2 inch rolled hose from shoulder height, rather than from the ground; reducing the weight of the rolled hose in the forcible-entry simulation from 80 to 60 pounds; and increasing the length of the rest period from 5 to 7-1/2 minutes.

The qualifying test became Exam 1162 with two additions and two modifications. A wall vault and hose pull were added. In addition, the 7-minute rest interval was converted to a 2-minute rest interval. Finally, instead of being a

qualifying test rated on a pass/fail basis, Exam 1162 became a competitive event, rated on the basis of a candidate's time to complete the series of evolutions.

The wall event originally proposed by the City was the 5 foot wall climb which formed the crucial part of Exam 3040 and had been the subject of extensive criticism in connection with the proceedings leading to the invalidation of that exam. Accordingly, the defendants not surprisingly acceded to plaintiff's demand that the height of the wall be reduced to 4-1/2 feet in order to conform more closely to the tasks of firefighters in climbing over obstructions in the course of responding to a fire. At the same time, defendants also acceded to plaintiff's demands that a somewhat duplicative walk through or over

tires, initially proposed as an additional test, be eliminated, in part at least because of plaintiff's argument that such a test, drawing on the experience of men in the military and in training for competitive athletic contests, would adversely impact women who lacked such a background.

The hose pull was added because it had been part of the set of job performance evaluators used in Philadelphia from which other essential elements of the qualifying test had been drawn, which had been eliminated only because of a misunderstanding on the part of a representative of the City's Department of Personnel as to the manner in which a hose pull was performed by firefighters in New York City. According to credible testimony of Dr. Phillip Siegel of the Department of Personnel, the hose pull

was left out of the qualifying exam because of his understanding that, when the task of pulling an uncharged length of hose to an upper story of a building over a hose roller attached to a window or roof was performed in New York City, it was done by two or more firefighters working together. When this misunderstanding was corrected, the hose pull using a hose roller was restored to the series of evolutions used in Philadelphia to measure job performance. While plaintiff correctly faults defendants for the somewhat ad hoc manner in which the hose roller subtest was added and argues that its adverse impact on women was not adequately studied, she does not offer any persuasive evidence to answer defendants' proof that the task does, in fact, reflect an important firefighting activity and calls upon physical capacities,

namely, upper body strength, required by a number of other firefighting tasks not otherwise tested in the battery of physical tests developed.[3] Moreover, plaintiff's complaints concerning the inadequacy of the job analysis performed by the City as a basis for adding the hose pull are considerably undercut by her acknowledgement that the subtest had its origin in the same series of job evaluators from which the other subtests reflecting a representative group of important job tasks were selected. Nor is the subtest as actually administered, after a familiarization program, of such complexity or difficulty as to justify fears

---

[3] For example, it appears undisputed that overhauling and pulling of ceilings form an important part of firefighting in New York City and call upon reserves of upper body strength not otherwise tested in Exam 1162.

that it tested learned behavior as opposed to a candidate's capacities to be trained to be a New York City firefighter.

The principal ground for contention between the parties relates to their differing approaches to an issue which first arose in connection with Exam 3040, namely, the issue as to the role of speed versus pacing in the performance of firefighting or, to put it another way, the need for anaerobic as opposed to aerobic energy in performing important firefighting tasks. On this issue, this Court had previously rejected the City's position that firefighting is an all-out sprint event performed at maximal heart rate, calling on reserves of [an]aerobic

energy in connection with Exam 3040,[4] and plaintiff's arguments that defendants failed to develop an adequate job analysis with respect to this issue carry considerable force.

In its opinion invalidating Exam 3040, this Court found:

> "Few jobs making large physical demands, least of all firefighting, are properly performed at maximum speed or at the limits of one's strength or endurance. As [plaintiff's] witnesses testified, what must be identified are not those who are strongest or fastest but, instead, those who, with the benefit of training in pacing or because of their natural capacities of endurance, can perform the punishing tasks of firefighting as they are actually required to be performed. According to these witnesses, fire-

---

[4] As previously noted, the facts that the qualifying exam included a 7-minute rest period and was rated on a pass/fail basis militate against any finding that the successful performance of plaintiff and other class members on the qualifying exam demonstrates that all-out speed as opposed to pacing is of the essence in firefighting.

fighting takes its toll, not as a result of failures of maximum strength or speed, even at critical moments, but rather through the physical demands extending over long periods of time which necessitate paced performance at less-than-maximum levels. This explanation makes sense of a number of otherwise puzzling features of firefighting, namely, the ability of firemen to continue to do their work competently over an entire working career and the results of physical testing of incumbent male firemen which show that on a variety of measures of maximum physical capacity firemen rate no better than the average American male. It also makes sense in terms of the recognized dangers of firefighting and their unpredictability which, as numerous witnesses testified, make hazardous in the extreme performance at top speed or at the limits of strength or capacity. Not only does maximal performance compromise the firefighter's ability to pace performance over the long periods of time during which physical demands are being made on the body, in addition, the very unpredictability of fire and the instability of burning structures call upon qualities of foresight, endurance, and pacing not examined by tests of maximum physical strength."

Berkman v. City of New York, supra, 536 F. Supp. at 212 (footnotes omitted).

Despite these findings and this Court's stated readiness to hear additional proof on the issue, defendants failed lamentably to establish a basis for the emphasis placed on maximal strength and speed in Exam 1162.

Three separate aspects of the physical portion of Exam 1162 are principally responsible for the test's emphasis on speed and anaerobic energy as the distinguishing characteristics of applicants meriting selection as a New York City firefighter: first, the notice to applicants that performance would be measured on time to complete the physical tests; second, the brevity of the test; and, third, the relatively narrow banding of the ranks from which appointments will be made. With respect to these three hotly contested aspects of the exam,

defendants performed an inadequate job analysis.

The principal basis for the exam's emphasis on speed was a questionnaire distributed to a small group of job experts on October 18, 1982. Since defendants had already, in June 1982, distributed a job-analysis questionnaire to approximately 2,000 firefighters which could have but did not address the key issue of pacing, and since the distribution of the October questionnaire followed rather than preceded the defendants' application to this Court in October 1982 for approval of Exam 1162 on a rank ordered time to complete basis, there is considerable substance to plaintiff's contention that the questionnaire was conceived of as an effort to rationalize a decision to emphasize speed and anaerobic energy, rather than as a neu-

tral inquiry to determine the role of pacing in the performance of actual firefighting tasks.

However, quite apart from the circumstances giving rise to the distribution of the questionnaire, its content hardly permits a conclusion that it constituted a neutral inquiry into the question of the pace at which firefighting is actually performed. Instead, the questions appear designed to lead respondents to answers that are supportive of defendants' oft-expressed contention that firefighting tasks are performed at top speed and at maximal energy levels. Thus, rather than ask the respondents to describe the pace (or even speed) at which various firefighting tasks are performed, the questionnaire describes the proposed new physical test for entry-level firefighters and then baldly asks

whether firefighters performing those tasks should complete them in the "fastest possible time," assuming only that they are not hindered by smoke or obstructions. Quite apart from the dangers of alerting respondents to the significance of their answers in terms of constructing a rating system for a new physical exam and the ambiguity of the phrase "fastest possible time" in terms of what factors, if any, other than smoke and obstructions should be taken into account in framing an answer, the leading nature of the question quite obviously suggests its answer, and not surprisingly the largest number of respondents answered "yes." Nevertheless, the obvious qualitative common sense of a number of the comments offered by those able to resist the invitation to respond affirmatively should have indicated to a neutral ob-

server both the ambiguity of the question and the need for further inquiry. Thus, among the responses from incumbent firefighters (on a questionnaire which did not even leave space for comments) were the following:

> "What is considered maximum possible speed? On the fire scene most is done walking, not running, for safety reasons"; "Hose stretching at fire operations is never performed at maximum speed (or to point of exhaustion) because the truly arduous task begins at the apartment door"; "If operating at maximum speed one would tend to get emotional and emotions could interfere with clear thought"; "[F]ire operations require speed, tempered with safety and corrections of actions. The terms fastest possible time and maximum possible speed seem to represent an overemphasis on speed."

Essentially the questionnaire responses which form the basis for the exam's emphasis on speed and anaerobic activity do no more than reveal the obvious, that <u>all else being equal</u> everyone would like fires to be extinguished and

victims extricated in the shortest time possible given the constraints of the job. However, it is precisely the question as to what the constraints of the job are that remained unexplored, with the exception of smoke and obstructions on stairs, together with the effect of those constraints on the speed with which firefighters actually do their work. The defendants' failure to explore in a neutral and above-board fashion this well-identified issue can only confirm this Court's earlier conclusion that a test that places undue emphasis on anaerobic energy and speed invidiously discriminates against women.

It is also the case that defendants failed to pay reasonable attention in constructing the test to the degree of emphasis placed on anaerobic energy by the manner in which the test was adminis-

tered. Essentially, the physical test consisted of two bouts of violent physical exercise which applicants were encouraged to perform at maximum speed (and which, if performed at a pace meriting selection according to the City's rating scheme, had to be performed in approximately 90 seconds) separated by an enforced 1 minute 40 second recuperation period. As plaintiff points out, the effect of inserting the rest period between two episodes of anaerobic exercise converts the test into the equivalent of "wind sprints" in which persons who exhaust themselves by a maximum expenditure of anaerobic energy on the engine company simulation are given a recovery period permitting them to rely again principally on anaerobic energy to complete the ladder company portion of the physical exam. Defendants appear never to have consid-

ered that a test taking the same time to administer, but filled completely with strenuous physical activity, might not only resemble more closely the level of physical activity actually required of firefighters at fires, but also might enforce pacing on the candidates, giving some reading of the candidate's aerobic energy. In structuring the test, defendants appear to have ignored not only this Court's prior findings concerning the role of aerobic energy in performing firefighting functions but also the recommendations of the same expert whose Philadelphia evaluators appear to have been the principal source of both the qualifying exam and Exam 1162 that the

test last 5 to 10 minutes without a recovery period.[5]

What has been said concerning the demands of firefighting should not be taken as a finding that firefighting is entirely an aerobic activity, making no demands for the expenditure from time to time of anaerobic energy at maximal oxygen levels, but only that defendants' exam, including in particular their proposed rating system, places too great an emphasis on the anaerobic capacities of candidates with too little regard for those aspects of firefighting that call

---

[5] Significantly, Chief Homer Bishop testified that firefighters need at least 5 to 10 minutes to get a one-room fire "under control," not counting additional time needed for overhauling and take-up operations. Other evidence established that the average time spent by firefighters at a "good job" is 40 to 60 minutes with a range from 25 minutes to 2 hours or, in unusual cases, 7 hours.

upon prudent, paced performance. Ironically, this finding is confirmed by defendants' criterion study which shows a lower correlation between raw scores achieved on the physical test and the performance of persons taking the test as evaluated by defendants' criterion measures than when a banding system was used, a subject discussed more at length below.

Finally, defendants appear to have paid singularly little attention to the emphasis placed on anaerobic energy by the narrow thirty-second banding of the scoring system selected and its extraordinary adverse impact on women. Indeed, the scoring system selected by the City, given the realities of the hiring process, appears to have virtually assured that no women would be appointed from the eligibility list created for

Exam 1162, despite the substantial number of women who passed the exam.

Based on past experience with other eligibility lists, defendants predict that they can reasonably expect to hire only some 2,800 firefighters from the total of 21,313 candidates who passed the exam. Again, based on past experience, the City predicts that offers of employment may be expected to be extended to approximately 6,500 applicants on the list in order to find the 2,800 candidates needed to fill the fire department's needs over the expected life of the eligibility list. To reach these top 6,500 candidates using the rank order proposed by the City, the City would get no farther down the list than to those candidates with a combined score on the written and physical tests of 94.5 or above. There are, however, 7,344 persons

on the current list with a score of 94.5 or above, and, of this group, only two are women. Accordingly, the City's rank ordering decision appears of crucial significance in determining the adverse impact of the physical exam on women.

This Court's earlier decision directed defendants in their preparation of a new exam to validate the exam in general accord with the Uniform Guidelines including the Guidelines' requirement that alternative measures with equal validity but with less adverse impact on the protected group be explored and used where available. 29 C.F.R. § 1607.14(B)(9). Accordingly, it is of particular significance that the alternative three-band scoring system proposed by plaintiff shows greater validity than the seven-band rating system proposed by the City. Use of the three-band rating

system accordingly appears required by defendants' own criterion measures[6] as well as by the failure of defendants in

---

[6] Plaintiff launches a considerable attack on the defendants' criterion measures, relying principally on her concerns that the "highly polarized" atmosphere prevailing within the department at the time the criterion studies were performed and the small sample of women available within the department to be evaluated render the results of the criterion studies suspect or worse. In ideal circumstances, certainly such concerns are well-founded: a more neutral atmosphere and a larger sample of female firefighters would, of course, be desirable to ensure reliability of the City's criterion studies. However, test preparation and the application of the Guidelines must go forward in the real world including situations such as that presented here where a charged and contentious atmosphere results from ongoing litigation and the limited number of female incumbents results from the relatively recent interest of women in firefighting work. What the Guidelines enjoin are reasonably competent efforts to prepare a non-discriminatory exam, not perfection. Plaintiff's arguments that the criterion studies were deliberately rigged to validate Exam 1162 are not borne out by the evidence of record.

their job analysis and test preparation to give due consideration to the demands made on aerobic energy in performing firefighting tasks. Finally, the selection of firefighters from among broader rating bands will serve to redress the undue emphasis placed on the physical portion of Exam 1162 as a result of the extraordinarily indiscriminate written portion of the exam.

There remains for consideration what steps, if any, should be taken because of the extraordinary fall off of interest on the part of women candidates who passed the written exam when called on to present themselves for physical testing. Had women passing the written exam presented themselves for the physical exam in the same proportion as their male counterparts who passed the written test, some 432 women would have taken the

physical exam as opposed to the 165 who actually did so. Here, even more obviously than was the case with respect to Exam 3040, it may be inferred that continued discrimination within the fire department in the form of highly publicized events leading to the termination of plaintiff and class member Gonzalez served to discourage women applicants from joining the fire department ranks. For this consequence of the fire department's continued discrimination against women some remedy appears appropriate.

While counsel for defendants has proposed offering the women deterred from taking exam 1162 a new chance to do so, a number of considerations weigh against this course of action. First, a substantial time lapse has occurred between the time the physical portion of Exam 1162 was administered to women and

now, so that age and lack of conditioning will undoubtedly continue to deter a substantial number of women once interested in taking the test from taking it at this time. Equally important, nothing defendants have done in the interim since September 1983 has served to allay the fears generated at the time that the fire department has failed to correct the intentional discrimination against women in its ranks which this Court found existed in its decision restoring plaintiff and class member Gonzalez to the ranks. Finally, no women who passed the written exam but did not take the physical have come forward to express an interest in this alternative. Accordingly, it seems more appropriate to consider the interests of plaintiff and the class she represents, namely, the group of approximately 38 who are currently women fire-

fighters, in fashioning a remedy that affords them colleagues who are committed to becoming firefighters and not easily deterred by the department's failure to eliminate discrimination in the ranks.

With these considerations in mind, it seems appropriate that, in the random selection process within each of the ranks to be established pursuant to this decision with respect to the physical portion of Exam 1162, each woman who passed the exam will be afforded a somewhat greater chance of selection than her male counterparts. This will reflect the likelihood that, had 432 women presented themselves to take the exam, the number of women passing the exam would presumably have increased correspondingly from 77 (out of 165) to 201 (out of 432). Assuming that the 124 additional women would have been spread over the ranks in

a pattern corresponding to the 77 women who actually took the exam and passed, each woman on the eligibility list should be afforded an increased opportunity over an equally ranked male to be selected in a ratio of 2.62 to 1.

The parties have apparently not explored the impact on women of using a three-band system with respect to the physical portion of Exam 1162 in combination with the results of the extraordinarily indiscriminate written exam. Since the written exam fails to make meaningful distinctions between various candidates' cognitive ability, it may be that rating the written exam on a pass/fail basis or in bands will result in a test with equal validity but less adverse impact on women. The parties are accordingly directed to address themselves to this issue and to settle an

order on ten days notice together with supporting briefs and, if necessary, additional evidence that is consistent with this opinion.

In summary, defendants have constructed and administered a physical examination that tests for the important physical capacities necessary to be a successful firefighter. However, the scoring system used by defendants places undue emphasis on certain physical capacities not required for the job. Accordingly, defendants are directed to develop a new scoring sytem, using the three-band division of the raw scores on the physical portion of Exam 1162 proposed by plaintiff in combination with the results of the written portion of the exam scored on a basis that has equal or greater validity than the present scoring system and the least adverse impact on women.

Selection of women passing Exam 1162 using this scoring system shall compensate for the impact of defendants' discriminatory conduct in deterring qualified women candidates in the manner outlined above.

The clerk is directed to mail a copy of the within to all parties.

SO ORDERED.

Dated: Brooklyn, New York
October 8, 1985

s/Charles P. Sifton
United States District Judge

UNITED STATES DISTRICT COURT
EASTERN DISTRICT OF NEW YORK

- - - - - - - - - - - - - - - - x

BRENDA BERKMAN, etc.,

Plaintiff,

- against -

CITY OF NEW YORK et alia,

Defendants.

- - - - - - - - - - - - - - - - x

CV-79-1813

ORDER

Upon the findings of fact and conclusions of law set forth in this Court's Amended Memorandum and Order dated October 8, 1985, it is hereby ORDERED:

1. Defendants, their attorneys, officials, agents, servants, employees, and attorneys, and their successors and all persons in active concert or participation with them or any of them (hereinafter collectively referred to as "defendants") are hereby permanently

enjoined from publishing or promulgating, sending notices of appointment to, or taking any steps to appoint, any persons to the job of entry-level firefighter on the basis of any eligibility list derived from Exam 1162, except pursuant to this order.

2. Defendants shall forthwith promulgate an eligibility list from the results of Exam 1162 and make appointments therefrom, as needed, in the manner set forth below:

(a) Defendants are directed to develop a new scoring system using the three-band division of the raw scores on the physical portion of Exam 1162 proposed by plaintiffs on June 20 and 21, 1985. This scoring system shall place in the top band ("Band A") candidates who received a 100 or 95 on the physical test; in

the second band ("Band B") candidates who received a 90, 85 or 80; and in the third band ("Band C") candidates who received a 75 or 70. Candidates in Band A shall be assigned a score of 95.4; candidates in Band B shall be assigned a score of 87.6; candidates in Band C shall be assigned a score of 73.6.

(b) A normal distribution of test scores for the written portion of Exam 1162 shall be computer generated and randomly assigned to all candidates who passed the written and physical portion of the examination. These randomly assigned scores shall be combined with the results of the three-band scoring of the physical portion of Exam 1162 described above.

(c) Selection from this new eligibility list shall occur in the following manner: Female applicants on the list shall be afforded an increased opportunity over equally ranked males on the list to be selected pursuant to a ratio of 2.62 to 1. This ratio shall be known as the compensation ratio. To apply this compensation ratio, the defendants shall divide the number of eligible men assigned a particular score on the list by the number of eligible women receiving the same score, times the compensation ratio of 2.62. The quotient of the calculation shall determine how many men receiving a particular score on the examination may be selected before the next woman is selected. This ratio shall be known as the

selection ratio. Where men and women receive the same score, a woman shall be selected first, and the number of men determined by the selection ratio shall be selected thereafter. This process shall be repeated until the score is exhausted, and the next score shall be treated in the same manner, provided there is [sic] also women receiving that score.

3. To ensure that the ratio of male to female candidates appointed as a result of Exam 1162 is the same as if a non-discriminatory scoring plan had been in operation since the defendants' use of the results of Exam 1162 began, the eligibility list described above shall be developed including the names of candidates already hired. All female candidates who would have been heretofore

appointed from the eligibility list created pursuant to this order, had such list been in effect at the time candidates were first appointed from Exam 1162, shall be appointed to fill the next available openings for appointment of an entry-level firefighter in the order in which their names appear on the eligibility list created pursuant to this order. In order to ensure that no male candidate previously appointed from the results of Exam 1162 pursuant to prior orders of this Court loses his position as a result of this order, all male candidates previously appointed based on the results of such examination shall be deemed to have been entitled to priority of appointment with respect to any higher ranked male candidate (but not with respect to any higher ranked female candidate) on the list created pursuant to this order.

4. The Court shall retain jurisdiction of this action to ensure compliance with this order, for consideration of such further interim or remedial relief as may be necessary and appropriate, and for consideration of any application by plaintiff for attorneys' fees and expenses and for costs, pursuant to 42 U.S.C. § 2000e-5(k).

The Clerk is directed to mail a copy of the within to all parties.

SO ORDERED.

Dated: Brooklyn, New York
February 14, 1986

s/ Charles P. Sifton
United States District Judge

Brenda BERKMAN, on behalf of herself and a class consisting of all similarly situated women, Plaintiff-Appellee,

- v. -

The CITY OF NEW YORK; Edward Koch, individually and as Mayor of the City of New York; New York City Fire Department; Augustus Beekman, individually and as Fire Commissioner of the City of New York; New York City Department of Personnel; Michael Nadeo, individually and as Director of Personnel of the City of New York; Thomas Roche, individually and as former Director of Personnel of the City of New York; Civil Service Commission of the City of New York, Defendants-Appellees.

Uniformed Firefighters Association Local 94, Defendant-Intervenor-Appellant.

No. 526. Docket 82-7654.

United States Court of Appeals, Second Circuit.

Argued Oct. 8, 1982.

Decided March 29, 1983.

Norman Eric Teitler, Rego Park, N.Y., for defendant-intervenor-appellant.

Laura Sager, New York City (Women's Rights Clinic of the Washington Square Legal Services, Inc., Robert L. King, Jeffrey N. Drummond, Debevoise &

Plimpton, New York City, on brief), for plaintiff-appellee.

Before FRIENDLY, NEWMAN, and KEARSE, Circuit Judges.

KEARSE, Circuit Judge.

This appeal by defendant-intervenor Uniformed Firefighters Association Local 94 ("UFA") questions the propriety of certain injunctive relief granted by the United States District Court for the Eastern District of New York, Charles P. Sifton, Judge, in favor of plaintiff Brenda Berkman and a class of women who sought to become firefighters in the New York City Fire Department, against defendants City of New York and certain individuals responsible for hiring firefighters for the City of New York (hereinafter collectively "the City"), to remedy unlawful employment discrimination on the basis of sex, in violation of Title VII of the Civil

Rights Act of 1964 ("Title VII"), 42 U.S.C. §§ 2000e to 2000e-17 (1976 & Supp. IV 1980). Berkman contended that the physical part of a 1978 Fire Department entrance examination called "Exam 3040" was not job-related and had a disparate impact on women. After a bench trial, the district court ruled, in an opinion reported at 536 F. Supp. 177 (1982), that the City's use of the physical portion of Exam 3040 violated Title VII. The court therefore enjoined further use of the eligibility list compiled pursuant to the exam ("Eligibility List 3040") "except upon a showing of compelling necessity," id. at 218, directed the City to develop a valid physical test, and ordered the City in the interim to appoint as entry-level firefighters up to 45 of those class members "who are found to be qualified for appointment and willing to be

appointed," id. (footnote omitted). The court directed the parties to attempt to agree upon procedures for determining which of the class members were qualified for appointment. Id. Thereafter, the City and Berkman agreed on an interim qualifying physical test to be administered to the class members, and the district court approved their agreement.

The present appeal is taken by UFA, an employee organization representing current firefighters, which was allowed to intervene below in order to participate in the remedy phases of the proceedings. Pursuant to a preargument agreement among UFA, Berkman, and the City, and approved by the Court,[1] the

---

[1] The preargument order embodying the stipulation provided, in pertinent part, as follows:

> IT IS HEREBY ORDERED that in accordance with the consent of the par-
> (Footnote continued)

issues on this appeal are limited to the propriety of the remedial measures or-

---

(Footnote 1 continued from previous page)
ties herein, the issues to be raised on this appeal are:

A. Issues to be Raised by Appellant

1. Defendant-intervenor appellant appeals from paragraph 4 of the District Court's Order of March 25, 1982 on the grounds that:

a. The remedy of ordering the City to hire up to 45 women from the class members found to be qualified for the job of firefighters, by means of a special qualifying procedure to be agreed upon by the parties or determined by the Court, constitutes affirmative relief rather than compliance relief.

b. The Court should have ordered the City either to give a special qualifying test to all men and women who failed Exam 3040 and to hire them on a rank order basis or to allow the women who failed Exam 3040 to take the next entry-level exam for firefighter and to be hired on a competitive basis with all other candidates for the position.

(Footnote continued)

dered by the district court.[2] For the reasons below, we affirm.

(Footnote 1 continued from previous page)

c. The Court erred in ordering the City to hire up to 45 qualified women because the number 45 is based on a calculation that included all women who took and passed the written test for Exam 3040, regardless of whether or not they took the physical exam.

2. Appellant appeals from the District Court's Order of August 3, 1982 on the grounds that:

a. The maximum qualifying time of 4 minutes 9 seconds was based on testing a group of firefighters that included persons over the age of 29, which is the maximum age for applying for the job.

b. The qualifying exam does not test for upper body strength in the manner or to the extent to which the record shows it is required for the job.

c. The qualifying test sets lower physical standards than the physical test of Exam 3040.

Although the City was party to the stipulation embodied in this order, it

(Footnote continued)

(Footnote 1 continued from previous page) has elected not to participate in the appeal either by brief or by oral argument.

2 Notwithstanding the stipulation and order limiting this appeal to issues regarding relief, see note 1 supra, UFA appears to press in its brief on appeal certain challenges to the district court's decision as to the merits of Berkman's claims. In light of the parties' limiting stipulation, we do not address UFA's arguments of the merits except to note in passing that perhaps its challenges on the merits were well foregone. Its major such challenge appears to be essentially that Title VII analysis is inapplicable to physical examinations. Thus, UFA states:

> A physical examination is an unbiased examination. An individual is either strong enough to pass the examination or he is not strong enough.

(UFA brief on appeal at 10.) UFA's premise ignores the law. If a physical test or criterion is not job-related and its application in employment decisions has a disparate impact on persons protected by Title VII, Title VII is violated. E.g. Dothard v. Rawlinson, 433 U.S. 321 (1977); Blake v. City of Los Angeles, 595 F.2d 1367, 1374-75 (9th Cir. 1979), cert.

(Footnote continued)

The history of the City's use and development of tests for the selection of its firefighters is fully set out in the district court's comprehensive opinion, 536 F. Supp. at 180-204, familiarity with which is assumed. At issue in the present case was the physical portion of Exam 3040 given by the City in 1978. The written portion of Exam 3040 had been taken in December 1977 by 24,758 men and 410 women. Nearly 98% of the men (or 24,252), and 95% of the women (or 389), passed this portion, and all who had done so were eligible to take the physical portion. Of the men so eligible, 74%, or 18,060, took the physical test. Of the eligible women, only

---

(Footnote 2 continued from previous page) denied, 446 U.S. 928 (1980); Officers for Justice v. Civil Service Commission, 395 F. Supp. 378, 382 & n.1 (N.D. Cal. 1975).

23%, or 88, took the physical test. Of the men who took the physical exam, 16,925 completed it and 7,847, or 46% of those who completed it, passed. Of the 88 women who took the physical exam, 79 completed it and none passed.

The present action was commenced by Berkman in 1979. In 1980 the district court certified a plaintiff class consisting of women who took the written portion of Exam 3040 and either took and failed the physical portion of Exam 3040 or were deterred from taking it as a result of sex discrimination by the City. Berkman thereafter moved for preliminary injunctive relief prohibiting the City pendente lite from hiring additional firefighters from Eligibility List 3040; this motion was denied after the City agreed to reserve 50 entry-level firefighter positions for members of the

plaintiff class in the event Berkman should prevail on the merits. In April 1981, UFA moved to intervene in the action, and its motion was granted to the extent of permitting it to participate at the remedy stage of the case, as well as in earlier proceedings having a bearing on the nature of the remedies to be granted.

After extensive discovery, a bench trial consuming several weeks was held between September and November 1981. The district court rendered its decision in an opinion dated March 4, 1982. See 536 F. Supp. 177. The court found, inter alia, that Berkman had established a prima facie case of discriminatory impact under both the Uniform Guidelines on Employee Selection Procedures ("Guidelines"), 29 C.F.R. § 1607.4(D) (1981), promulgated by the Equal Employment

Opportunity Commission ("EEOC"), and the statistical test stated by the Supreme Court in Castaneda v. Partida, 430 U.S. 482, 497 n.17 (1977). 536 F. Supp. at 205-06. The court found that the City had failed to rebut plaintiff's case since it did not establish that the physical portion of Exam 3040 was job-related. The City had failed to produce either "data showing that the content of the selection procedure [was] representative of important aspects of performance on the job for which the candidates [were] to be evaluated," 29 C.F.R. § 1607.5(B) (sometimes referred to as "content validation"), 536 F. Supp. at 206-07, or "empirical data demonstrating that the selection procedure [was] predictive of or significantly correlated with important elements of job performance," 29 C.F.R. § 1607.5(B) (called

"criterion validation"). 536 F. Supp. at 208. On the basis of these findings the district judge concluded that the physical portion of Exam 3040 violated Title VII, and he granted the injunctive relief that is the subject of this appeal in orders dated March 25, 1982 ("March Order"), and August 3, 1982 ("August Order").

A. The March Order

The March Order granted immediate, long-term, and interim relief. Effective immediately, the court permanently enjoined further use of Eligibility List 3040 except upon a showing of compelling necessity. As long-term relief, the court directed the City to begin preparation of new, properly validated selection procedures that would have the least adverse impact on women. For the interim, the court ordered the

City to reserve entry-level firefighter positions for the hiring of up to 45 members of the plaintiff class who remained interested and were adequately qualified.

The court arrived at the number 45 with respect to interim hiring as follows. It found that some women who had passed the written portion of Exam 3040 had been deterred from taking the physical portion by pre-test publicity suggesting that no woman could pass the physical test.[3] 536 F. Supp. at 217.

---

[3] Prior to administering Exam 3040, the City had attempted to recruit women to take the test and offered a booklet to familiarize candidates with each component of the physical test and with preparatory exercises. City firemen conducted informal training sessions privately for a fee. 536 F. Supp. at 200. One result of these training programs was publicity predicting that no woman could pass the physical test. For example, the lead sentence of one New York Daily News article reported that "[a] practice test by two dozen

(Footnote continued)

The district court correctly concluded that a "sex-neutral exam would have been followed by an equal decline (26%) in interest between both groups." Id. The court reasoned that absent the actual and anticipated disparate impact of the physical test, the decline in interest among women who had passed the written portion could reasonably have been expected to be proportionately no greater than the decline in interest among the men who had passed the written portion. Since 74% of the men who had passed the written portion of Exam 3040 took the physical por-

---

(Footnote 3 continued from previous page) would-be women firefighters through the Fire Department's training course . . . showed that the physical test seems to rule out women." New York Daily News, Dec. 11, 1977, at 4. There was trial testimony from several members of the plaintiff class as to the discouraging effect of the publicity.

ion, the court estimated that 74% of the 89 women who passed the written portion, r 288, would have taken a sex-neutral hysical test. Then, noting that 16% of he men who had taken the physical test ad been called up for appointment,[4] and aking 16% of 288, the court arrived at a figure of close to 45 women who one ould expect to have been called up under sex-neutral exam." Id.

The March Order directed the arties to attempt to agree upon a proce- ure for determining which members, "if ny," of the plaintiff class were ade-

---

[4] As of the date of the court's decision, 2,666 men, or 16% of the 16,925 who completed the physical exam, had been called up for appointment. See note 13 infra. Subsequently the City obtained permission from the court to appoint additional firefighters from Eligibility List 3040, and the number of places to be reserved for women was also increased.

quately qualified to become entry-level firefighters and hence to be appointed to the reserved positions. Paragraph 4 of the order provided, in pertinent part, as follows:

> 4. In order to determine which, if any, members of the plaintiff class are entitled to further remedial relief, the parties shall forthwith determine which members of the plaintiff class (a) continue to be interested in the position of firefighter, and (b) are adequately qualified for that position. In this connection plaintiff shall, on or before April 2, 1982, serve and file on 3 days' notice to all parties its proposed form of notice to the class, directing members to notify plaintiff's counsel of their intent to qualify for appointment within a reasonable period of time to be specified in the notice. In the event that the parties are unable to agree expeditiously upon procedures for the determination of which class members responding to such notice are adequately qualified, then the procedures for determination of whether interested class members are adequately qualified for the position shall be, on notice to the parties, established by the Court. In the event that more than 45 members of the plaintiff class are found to be interested in and adequately qualified for appointment

to the position of entry-level firefighter, notices of appointment shall be sent to 45 of those qualifying, selected by lot. In the event that 45 or fewer members of the plaintiff class are found to be interested in and adequately qualified for appointment, notices of appointment shall be sent to all those qualifying.

The Court refused to order that the City affirmatively recruit women to fill the 45 reserved places, finding that Berkman had failed to prove either a pattern of significant prior discrimination or that the City's discrimination against women had been intentional. 536 F. Supp. at 217.

3. The August Order

The August Order approved a stipulation between Berkman and the City agreeing on an interim physical test to be used to determine which members of the plaintiff class were "adequately qualified" to be firefighters and hence could be appointed to the reserved positions.

The events leading to the August Order began in the wake of the March Order. Berkman and the City had different ideas as to the most appropriate procedure for determining which members of the plaintiff class were adequately qualified to become firefighters. The City favored development of a new physical exam, whereas Berkman sought a training program in which interested class members would participate and have their fitness to be firefighters evaluated solely on the basis of their performance in that program. Each side proceeded to develop its own proposal.

The City's Department of Personnel undertook a review of the literature regarding firefighter testing, existing analyses of the New York City firefighter's job, and the district court's opinion in the present case.

Eventually, after consultation with members of the Fire Department, UFA, and outside consultants, the City devised a two-part physical exam, as described in the margin.[5] This exam was itself tested

---

[5] The description of the interim test devised by the City was as follows:

> Part I--Engine Simulation Subtest
> A. Hose Stretch: Candidate holds one length of 3-1/2" hose weighing 80 pounds and stretches it 145 feet.
> B. Hose Carry: Candidate picks up a folded 2-1/2" hose weighing 46 pounds and carries it from the building entrance to the fifth floor by way of the stairs.
> --Five Minute Rest--[see note 6 infra]
> Part II--Ladder Simulation Subtest
> C. Ladder Raise: Candidate raises a 20 foot ladder that weighs 58 pounds and is lying on the ground from the horizontal to the vertical position.
> D. Ladder and Stair Climb: Candidate climbs a pre-set supported ladder up to the second story and then enters the building through an open window, picks up an 8 pound mallet and a halligan tool, and runs up to the fifth floor.
> E. Forcible Entry: On the fifth floor, candidate hits a rolled 3-

(Footnote continued)

on a group composed of 37 incumbent firefighters, selected to be representative of members of the Fire Department with respect to assignment, age, and years on the force. The testing was supervised by 20 superior officers, who were instructed as to what constituted acceptable and unacceptable performance on each subtest. A score of 4:09 minutes for performance of all the tests was eventually set as the cutoff passing score.[6] This was

---

(Footnote 5 continued from previous page)
1/2" hose weighing 80 pounds that is placed at one end of a table and drives it 12-1/2 feet to the other end of the table.
F. Rescue Drag: Candidate drags a 145 pound articulated dummy from a position on the fifth floor to a door leading to the stairs.

In performing all tasks, candidates were to wear a 24-pound "Scott Pak" and full firefighter uniform.

[6] A preliminary group of eight firefighters was tested on May 11, 1982. After the first four had completed the test, the officer-observers recom-
(Footnote continued)

erived from the distribution of the test cores of the incumbent test-testers, the ecommendations of the Fire Department fficers who had supervised the incum- ents in the test, and the recommenda- ions of a psychometrician retained as a onsultant by the Personnel Department.

In the meantime, Berkman, who new the City was working on an interim est but did not know its contents, con- inued to press for a physical training rogram. The City met with Berkman's ounsel several times to discuss the

---

Footnote 6 continued from previous page)
mended that the rest period between Parts I and II be shortened from one hour to five minutes, and the recommendation was adopted immediately. The remaining four firefighters who took the test on May 11, and the main group of 29 firefighters who took the test on May 14, performed the various subtests on the revised schedule. The performances of the first four tryouts were not considered in the computation of the recommended cutoff score.

nature and cost of Berkman's proposal, but did not disclose any information regarding its interim test prior to May 21, when the test was finalized. It did not invite Berkman to participate in the test development.

On May 21, in an effort to have interim hiring qualification procedures commenced without further delay, Berkman moved to have the district court approve her proposed physical training program. The City quickly cross-moved to have its proposed interim test approved. On May 26, the district court held a hearing on the applications, at which Berkman, the City, and UFA were represented.

The City urged approval of its proposed interim test, arguing that it was job-related and would permit the City to select persons who would be able to perform firefighting duties in New York

City. Berkman adhered to her preference for a training program, citing her lack of participation in the development of the City's proposed test, her consequent uncertainty as to the validity of the test, and her desire to end delay in gaining the interim relief granted by the March Order.

The court expressed concern over the secrecy with which the City's test had been prepared, stated its preference for a qualifying procedure based on "observations over a substantial training period" (May 26, 1982 Tr. at 32), and questioned the fairness of asking class members to take a physical test on short notice and without training procedures. The court reserved decision as to each side's proposal and encouraged the parties to attempt to compromise.

At the end of this hearing, UFA, which had offered no proposal of its own and had submitted no papers, testimony, or argument either for or against either proposed qualifying procedure, appeared to favor the City's proposed test. When the court indicated that it would reserve decision and asked counsel to return the next day, counsel for UFA suggested that perhaps a decision by the court could be obviated if Berkman's experts were to review the City's proposed exam and find it acceptable:

> MR. TEITLER [counsel for UFA]: Your Honor, before a determination is made by the bench, the plaintiffs have indicated that they haven't had their experts have time to go over the exam broached by the City.
>
> Wouldn't it be more advantageous to see if between us, if their experts review the exam proposed by the City and find it acceptable that it might give a little more leeway as to what the result might be?
>
> THE COURT: I don't have any problem at all with you discussing

it amongst yourselves and coming up with any resolution you can. I would encourage it . . . .

. . . If you can take the problem out of my hands, I would be happy to have you do it.

(Id. at 60-61.)

Compromise was not quickly forthcoming, however, and the parties so notified the court. Accordingly, on June 3, 1982, the court ruled orally, in essence, that it would approve the City's proposed interim test if it could be validated.[7] It ordered the City (1) to

---

[7] The court expressed its concerns with respect to validation and the need for expedition as follows:

> I certainly do not propose to approve this proposed qualifying physical examination that has been presented to me by the City without an opportunity for [plaintiff] to be heard, and also an opportunity for [plaintiff] to find out more about circumstances under which it was prepared, administered, and what its impact, validity and fairness may be, but I am also impressed as I was at the time I

(Footnote continued)

prepare to administer the proposed exam

to interested members of the plaintiff

(Footnote 7 continued from previous page)
asked you to proceed expeditiously to come to some resolution of this qualifying exam, with the need to get started in some direction, and I have been presented with a quite concrete proposal for a qualifying physical examination which is along the lines which during the trial of the matter seemed to be the kind of work based examination, which the plaintiff thought most appropriate.

. . . .

. . . [I]f this examination has been administered to people of substantial experience as incumbents in the Fire Department, there is a real question in my mind as to the degree that their times are a factor of their skill and training, and I am also concerned that the examination appears to, from the papers that have been presented to me, appears to have been prepared and administered with little attention as far as I can see to equally available alternatives with a less adverse impact on women, and without any effort to determine how serious, if it's serious at all, the adverse impact on women of this examination might be . . . .

(June 3, 1982 Tr. at 4-5, 6.)

class in September, (2) to appear in court in early August for a hearing on the validity of the proposed exam, (3) to disclose to the plaintiff class all information on the validity, fairness, and adverse impact of the exam, and (4) to conduct, beginning on June 7, familiarization and training programs for members of the plaintiff class who wished to take the interim exam.[8] Berkman and the City were directed to cooperate in developing the familiarization and training programs. This order was embodied in a

---

[8] The court declined to adopt Berkman's proposal for qualification solely by means of a training program, stating that that proposal would be more appropriate for "a situation in which affirmative action on the part of the City was required because of intentional discrimination in the past." (June 3, 1982 Tr. at 9.) In its ruling on the merits the court had declined to find intentional discrimination. 536 F. Supp. at 217.

written order filed on June 9 ("June 9 Order"). UFA immediately appealed.[9]

Berkman and the City nevertheless continued to negotiate, and eventually they agreed upon a qualifying physical test that differed only in minor respects from the interim test proposed by the City in May. The agreement was embodied in a stipulation dated July 12, 1982, which described the physical test as set forth in the margin.[10]

---

[9] UFA had previously appealed from the March Order; that appeal had been withdrawn, without prejudice, by stipulation dated June 7, 1982.

[10] The interim test agreed to by plaintiff and the City was as follows:

> The candidate will wear full turnout gear, including a turnout coat, gloves, boots, Scott Air Pak and helmet.
>
> PART I-Engine Simulation
>
> 1. Hose Stretch. Candidate holds one length of 3-1/2" hose weighing 80 pounds and stretches it 145 feet
> 2. Hose carry. Candidate

(Footnote continued)

UFA refused to join in the stipulation, however, and the City moved

---

(Footnote 10 continued from previous page)
transfers one folded length of 2-1/2" hose, weighing 46 pounds, from shoulder-height stand to shoulder, and carries it from the entrance of Building 1 to the fifth floor.

Seven and one-half minute rest.

PART II-Ladder Simulation (Rescue)

3. Ladder raise. Candidate raises a 20-foot ladder that weighs 58 pounds from the ground to a vertical position.

4. Ladder and stair climb. Candidate climbs a pre-set supported ladder up to the second story and enters the building through an open window; picks up a bar weigh ing [sic] 16 pounds, and climbs to the fifth floor.

5. Forcible entry. On the fifth floor, candidate, using a sledge type hammer weighing 8 pounds, hits a rolled 3-1/2" hose, not to exceed 60 pounds in weight, the length of a 12'6" table; the resistance factor of the hose on the table shall be no greater than nine kilograms of horizontal force as measured on a device to be supplied by Dr. William McArdle.

6. Simulated rescue. Candidate drags a 145-pound articulated

(Footnote continued)

to have the district court compel UFA to sign. At a hearing on August 3, the district court stated that it had no authority to order a party to sign a stipulation but that it would construe the City's application as a request that UFA "show cause why, if there is any reason, why this qualifying test shouldn't be approved." (Aug. 3, 1982 Tr. at 3.) The court then put that question to UFA's counsel:

(Footnote 10 continued from previous page)
dummy along a marked path on the fifth floor.

SCORING

All tasks are performed in sequence.

Maximum satisfactory time for performance of the test, not counting the rest period, shall be four minutes, nine seconds.

THE COURT: . . .

. . . .

Mr. Teitler[,] [p]erhaps you could tell me what your position is?

MR. TEITLER: My position is very simple, your Honor.

We are in the midst of appealing Court Order dated June 9th, and we feel if we sign any stipulation, it would prejudice our rights under the appeal.

THE COURT: All right. Well, I'll note your opposition, your note that you don't consent to the relief that's being requested and otherwise approve the qualifying test . . . .

(Id. (emphasis added).) UFA's counsel proffered no other objection and made no other statements.

After further discussion of minor details, the court so-ordered the stipulation signed by Berkman and the

City. This was the August Order that is one of the subjects of this appeal.[11]

After filing its notice of appeal from the August Order, UFA applied to the district court for a stay of the holding of the interim physical test pending the outcome of its appeal. In support of its application, UFA submitted affidavits by its counsel and its president asserting, inter alia, that the interim physical test approved by the court on August 3 was inadequate to establish whether candidates were qualified to be firefighters.[12] At a hearing on

---

[11] On August 5, 1982, UFA withdrew, without prejudice, its appeal from the June 9 Order and on August 27 filed its present appeal, which seeks review of both the August Order and the March Order. See note 9 supra.

[12] The affidavit of UFA's president stated that "[t]he thrust of [UFA's] appeal is that the bulwark of this department has been built on esprit de corps." (Affidavit of Nicholas

(Footnote continued)

September 2, the court denied UFA's application, stating as follows:

(Footnote 12 continued from previous page)
Mancuso, sworn to Sept. 1, 1982, at 5), and was filled with predictions of the erosion of firefighter morale and public confidence:

> Lowering the standards to allow the members of the class in will cause irreparable harm in opening the floodgates and enable future candidates of both sexes to enter the department with substandard qualifications. The high caliber of the men who have been members of this department and have epitomized to the world at large that the members of the New York City Fire Department are the best there is [sic], will hereinafter be undermined. This will have a direct effect on the conceptions of the public and will shake the faith of the community at large. No longer will the Fire Department symbolize unflinching devotion to duty and the ideal of putting ones [sic] life before that of the publics [sic] will fade in the light of this decision.
>
> The unit as an entity will be weakened and discipline thrown to the wind. Firefighters will have the seeds of doubt sown in their minds and fear will gnaw at their beings, wondering if those behind them will have the strength and

(Footnote continued)

THE COURT: You are raising questions about [this] qualifying examination, which I asked someone from your office whether they intended to raise back, as I recall, in the beginning of this month [sic: last month?], at a time when I said I wanted to hear those, so that these people who are preparing to take this test would not spend a month o[r] more preparing to take the test which was subject to complaints, and the only answer I got from your representative there was that simply because they felt it would be inconsistent with the position that they were taking on the [appeal] that they were not going to consent to the test, test form. I didn't get a pray [sic] of an explanation of any deficiencies that you found in this test. As a result people have gone out and literally, I assume, sweated through a month in preparation, and I really wonder, that you come in on the eve of the time when these people are going to go out and match their physical capacities with a test of this sort, and start to raise questions about the adequacy of the exam. . . .

. . . .

(Footnote 12 continued from previous page)
fortitude to shoulder the burdens and responsibility for the lives of each-other [sic].

(Id. at 2-3)

> . . . I'm going to deny the request for a stay pending the appeal. I find I can only conclude that it's being brought for the purposes of delay.

(Sept. 2, 1982 Tr. at 11, 13 (emphasis added).) A similar motion to this Court was likewise denied.

II. DISCUSSION

UFA challenges both the March Order and the August Order on several grounds. It challenges the March Order's requirement that places be reserved for up to 45 women who may be hired as firefighters on the grounds (1) that the order granted "affirmative" relief, which was improper in the absence of a finding of intentional discrimination or an egregious past history of discrimination, (2) that this relief discriminated against men who failed the physical portion of Exam 3040, and (3) that the number 45 was too high. UFA contends that the August

Order should not have approved the stipulated interim physical test because that test (1) improperly set a passing grade by reference to the performances of persons too old to qualify as entry-level firefighters, (2) did not provide an adequate test of upper body strength, and (3) lowered the standards, as reflected in Exam 3040, for becoming a New York City firefighter, thereby endangering public safety. We find no merit in any of UFA's contentions.

A. The Propriety of Reserving 45 Places for Interim Hiring

Once a violation of Title VII has been established, the district court has broad, although not unlimited, power to fashion the relief it believes appropriate. The bounds of the court's discretion are set by the purposes of Title VII, which are to prevent discrimination and achieve equal employment opportunity

in the future, see International Brotherhood of Teamsters v. United States, 431 U.S. 324, 364 (1977); Albemarle Paper Co. v. Moody, 422 U.S. 405, 417 (1975); Griggs v. Duke Power Co., 401 U.S. 424, 429-30 (1971), and to make whole the victims of past discriminatory practices, see, e.g., International Brotherhood of Teamsters v. United States, supra; Franks v. Bowman Transportation Co., 424 U.S. 747, 764 (1976); Albemarle Paper Co. v. Moody, supra, 422 U.S. at 418. Our inquiry with respect to the March Order is whether, in requiring the City to hire up to 45 members of the plaintiff class who are found to be qualified, the court abused its discretion. Id. at 421-22; Association Against Discrimination in Employment v. City of Bridgeport, 647 F.2d 256, 279 (2d Cir. 1981), cert. denied, 455 U.S. 988 (1982) ("ADE v. City

of Bridgeport"); cf. Kirkland v. New York State Department of Correctional Services, 520 F.2d 420, 429-30 (2d Cir. 1975) (employment discrimination action brought under the Fifth and Fourteenth Amendments to the Constitution and 42 U.S.C. §§ 1981 and 1983), cert. denied, 429 U.S. 823 (1976). We conclude that it did not.

1. The Number 45

We begin by considering the propriety of the court's selection of the number 45 for interim hiring, since that question has relevance not only to UFA's contention that that number was improperly inflated by the court's consideration of women who did not take the discriminatory physical test, but also to UFA's contention that this provision improperly granted affirmative relief in the nature of a quota.

We see no impropriety in the court's selection of the number 45. The March Order was linked precisely to the disparate impact that the court had found in the physical portion of Exam 3040. The court found not only that women had failed because the test was discriminatory, but also that women had been deterred from even taking the test because of the advance publicity predicting that no woman could pass. Those who have been deterred by a discriminatory practice from applying for employment are as much victims of discrimination as are actual applicants whom the practice has caused to be rejected. E.g., Dothard v. Rawlinson, 433 U.S. 321, 330 (1977); International Brotherhood of Teamsters v. United States, supra, 431 U.S. at 365-66 ("When a person's desire for a job is not translated into a formal application solely

because of his unwillingness to engage in a futile gesture he is as much a victim of discrimination as is he who goes through the motions of submitting an application"); ADE v. City of Bridgeport, supra, 647 F.2d at 281-82 & n.23. It was thus well within the court's discretion to fashion interim relief that took into account the entire discernible impact of the City's discriminatory test, and not just its most obvious effect, i.e., that all women who actually took the test failed.

It was also reasonable for the court to estimate the number of women deterred from taking the discriminatory physical test by referring to the percentage of eligible men who took the test, rather than seeking an actual deterrence census of the 301 eligible women who elected not to take the physical

portion of Exam 3040. Accordingly, we find no abuse of discretion in the district court's selection of the number 45 with respect to the interim hiring requirement imposed on the City.[13]

---

[13] In setting the interim number at 45, the court did not purport to find the exact number of women who would have been called up for appointment absent discrimination, but found only that 45 was "close to" that number. In chart form, the statistics discussed in Part I.A above are as follows:

Men			Women	
Number	% of Prior Category		Number	% of Prior Category
24,252	98%	Passed Written	389	95%
18,060	74%	Took Physical	88	23%
16,925	94%	Completed Physical	79	90%
7,847	46%	Passed Physical	0	0%

The court's own calculation of the number of women who would have been called up, see text accompanying note

(Footnote continued)

2. The Nature of the Relief

Title VII remedies ordinarily

(Footnote 13 continued from previous page)
4 supra, would have yielded a total of 46 (74% of 389 = 288, 16% of 288 = 46).

We find one slight flaw in the district court's calculation, which we regard as harmless. The court arrived at its 16% figure by dividing 2,666, the number of men by the[n] called up for appointment from Eligibility List 3040, by 16,925, which the court mistakenly stated was the number of men who presented themselves to take the physical exam. 536 F. Supp. at 217. The actual number of men who so presented themselves was 18,060 and 16,925 represented the number of men who completed the exam. Id. at 204. Thus, the court's calculation did not take into account the likelihood that even without the test's sex-discriminatory impact, a certain number of women would normally be expected not to complete the exam, and hence not qualify for appointment. The more appropriate calculation would have reduced the number 288 (i.e., 74% of women who passed written test) to 94% of that figure (percentage of men taking but not completing physical test) to reach 271 as the number of women who could have been predicted to complete a sex-neutral exam. Applying the 16% call-up figure to 271, one would arrive at 43 as the number of

(Footnote continued)

may be described as compliance relief, compensatory relief, or affirmative relief. ADE v. City of Bridgeport, supra, 647 F.2d at 278; Guardians Association of the New York City Policy Department v. Civil Service Commission, 630 F.2d 79, 108 (2d Cir. 1980), cert. denied, 452 U.S. 940 (1981) ("Guardians IV"). The three types of relief differ to some extent in purpose, effect, and justification.

Compliance relief is designed to erase the discriminatory effect of the challenged practice and to assure compliance with Title VII in the future. Such relief may include prohibiting the use of

---

(Footnote 13 continued from previous page) women to be called up. Given the fact that the court was not striving for a precise quantification, as discussed in the first paragraph of this footnote, we regard this slight variance as immaterial.

an invalid test or criterion, restricting appointments from an eligibility list compiled by reference to the results of an invalid test, ordering that new and valid selection procedures be adopted, and authorizing interim hiring that does not have a disparate impact on any group protected by Title VII. An interim hiring order does not have a disparate impact on any such group when (a) it orders the hiring of members of the plaintiff class whom the court has found to be victims of the defendant's discrimination, and (b) it calculates the number of victims to be so hired -- in relation to the total number of interim hirees -- by reference to the percentage the victims constituted of the total applicant pool. Such an order does not go beyond the simple elimination of the disparate impact of the practice found to be dis-

criminatory and is properly regarded as compliance relief. Compliance relief is appropriate whenever a Title VII violation has been found, irrespective of any history of prior discriminatory practices or the intent of the defendant. See ADE v. City of Bridgeport, supra, 647 F.2d at 278; Guardians IV, supra, 630 F.2d at 108 & n. 25.

Compensatory relief is designed to "make whole" the victims of the defendant's discrimination. To the extent that an order requires the hiring of a member of the plaintiff class -- i.e., a victim of the discrimination -- it constitutes both compliance relief and compensatory relief. Compensatory relief may also be granted in the form of backpay, payment of the value of past fringe benefits, and retroactive seniority. These forms of relief are generally

appropriate under the same circumstances as compliance relief. See ADE v. City of Bridgeport, supra, 647 F.2d at 278-80.

Affirmative relief is that designed principally to remedy the effects of discrimination that may not be cured by the granting of compliance or compensatory relief. It may include the setting of long-term hiring targets or the imposition of a requirement that the defendant actively recruit or train members of the Title VII-protected group. Such relief may be required where, for example, the defendant has intentionally or egregiously engaged in a practice of discrimination that is likely to have discouraged members of the protected group from becoming members of the applicant pool at any stage. Affirmative relief may also include interim hiring relief that is extended to persons other

than members of the plaintiff class and in proportions exceeding the ratio of plaintiff class members to the total applicant pool. See id. at 278, 282; Guardians IV, supra, 630 F.2d at 108-09.

Affirmative relief is normally justified only if the defendant's discrimination has been intentional, see, e.g., id. at 109, or there has been a long-continued pattern of egregious discrimination, see id.; ADE v. City of Bridgeport, supra, 647 F.2d at 286; Rios v. Enterprise Association Steamfitters Local 638, 501 F.2d 622, 631 (2d Cir. 1974). We have indicated, however, that affirmative relief for Title VII-type violations may be upheld even in the absence of these factors if it is limited to an interim stage and does not require the hiring of persons who are not members of the plaintiff class. Thus, in Kirk-

land v. New York State Department of Correctional Services, supra, in which the members of the plaintiff class comprised approximately 10% of the applicants who took a discriminatory promotional exam, we reviewed a provision that forbade interim promotions without court approval and required that at least one out of every four persons so promoted be a member of the plaintiff class. Although we struck down other provisions of the court's order that imposed permanent quota restrictions, we upheld this interim provision, stating as follows:

> Since this portion of the decree is interim in nature, does not mandate the making of any promotions, does not disregard an existing civil service eligibility list, and since its benefits are limited to the members of the plaintiff class, we affirm it as not being an abuse of the District Court's discretion.

520 F.2d at 429-30; see also Guardians IV, supra, 630 F.2d at 109 (hiring order

to assure Title VII compliance on interim basis may properly "select[] from among adequately qualified applicants . . . according to some appropriately non-compensatory ratio, see, e.g., Kirkland, supra, 520 F.2d at 429-30").

Within this framework, we think it plain that in the present case the court's order that up to 45 members of the plaintiff class be hired, if found qualified, constituted compliance relief rather than affirmative relief.[14] As discussed in the previous section, the number 45 was determined strictly by the court's estimate of the approximate number of women who had been the victims of

---

[14] The interim hiring provision has compensatory features as well since victims who are able to pass the interim test will be hired, and the City has agreed to pay them backpay and benefits with respect to a period prior to their actual appointment.

the City's discriminatory use of Exam 3040. Since only women who had passed the written portion of Exam 3040 were made eligible to take the interim test, the relief focused solely on the pool of applicants as it existed just prior to the discriminatory phase of Exam 3040.

The March Order neither imposed a quota -- i.e., the relatively permanent use of a specified hiring ratio -- nor established a hiring goal for the City. The order imposed no numerical long-range hiring requirements whatever; the only long-range relief ordered was the development of properly validated selection procedures that would not have a disparate impact on women. Further, the interim number 45 represented simply the maximum number of the plaintiff class members that the City would be required to hire. If more than 45 class members

took and passed the interim test, only 45 were to be hired. Nor was the City ordered to take any affirmative steps to attract women to the Fire Department. If fewer than 45 members of the class remained interested in taking the interim physical test, the City was not required to seek out others. If fewer than 45 passed the interim test, the City was to hire only the lesser number. If no women passed the interim test, the order did not require that any be hired.

Thus, the interim hiring provisions of the March Order simply provided an opportunity for the victims of the discrimination[15] to qualify by means of a

---

[15] We find no merit in UFA's contention that the provision for an interim test for interested members of the plaintiff class should be set aside unless it also is extended to men who failed the physical portion of Exam 3040. The provision was designed to remedy discrimination in violation of Title

(Footnote continued)

sex-neutral interim physical test. The district court properly regarded this as compliance relief, 536 F. Supp. at 216-17, which was justified by its finding of a Title VII violation, and we find in it no abuse of the court's discretion.

B. Approval of the Interim Test

In enacting Title VII, Congress expressed a preference for achieving compliance by voluntary means. See, e.g., Alexander v. Gardner-Denver Co., 415 U.S. 36, 44 (1974); United States v. City of Alexandria, 614 F.2d 1358, 1362 (5th Cir. 1980); Patterson v. Newspaper &

(Footnote 15 continued from previous page)
VII. There was no showing that the test had had any disparate impact on any group of men on the basis of their gender. Accordingly, there was no basis for awarding men, as well as women this interim Title VII relief. See, e.g., Patterson v. Newspaper & Mail Deliverers' Union, 514 F.2d 767, 772-73 (2d Cir. 1975); cert. denied, 427 U.S. 911 (1976).

Mail Deliverers' Union, 514 F.2d 767, 771 (2d Cir. 1975), cert. denied, 427 U.S. 911 (1976). In general, therefore, when the parties to the litigation have agreed on a means of achieving compliance, the district court should normally approve the settlement unless it contains provisions that are unreasonable, unlawful, or against public policy. On appeal, the district court's approval of a settlement should be upheld unless it constituted an abuse of discretion.

We find no abuse of discretion by the district court in entering the August Order approving the agreement between Berkman and the City to the terms of the interim qualifying test. We note, to begin with, that when the City commenced, following the March Order, to design an interim test, it appears to have done so with a proper concern for

public safety. The record indicates that the City sought expert advice throughout, consulted the Fire Department repeatedly, and received and implemented advice from UFA.

Moreover, prior to entry of the August Order, UFA never suggested that either the agreed-upon test or its City-prepared predecessor was in any way inadequate. UFA plainly had no quarrel with the adequacy of the proposed interim test developed by the City prior to May 26, 1982. UFA came forward with no criticism of that test, and indeed, indicated its own approval by suggesting that if Berkman's experts would review and could approve it, a court decision would be unnecessary. There appears to be little difference between that initially proposed test and the one that was eventually agreed on by Berkman and the City,

compare note 5 supra with note 10 supra; and when the court asked UFA why the agreed-upon test should not be approved, UFA gave no indication that it thought the test was inadequate. It stated only that it opposed the test because it had an appeal pending from the court's June 9 Order. The court stated it took this to mean that UFA challenged the nature of relief that had been granted but that it "otherwise approve[d] the qualifying test." (Aug. 3, 1982 Tr. at 3.) UFA gave no indication that this was not an accurate interpretation of its position. Thus, the district court was not presented with any basis for believing that it should not approve the interim test on grounds of inadequacy.

Nor do the particular criticisms of the interim test voiced by UFA on appeal suggest to us that it would

have been unreasonable for the district court to approve the test if UFA had timely voiced these criticisms below. We are unpersuaded by the complaint that incumbent firefighters over the age of 29 -- the maximum age for entrance into the Fire Department -- were used to test the test and that their scores were used to derive the passing grade of 4:09 minutes. To be sure, the performance times of the incumbent test-testers indicated that those over the age of 29 generally performed less swiftly than those who were younger. Two countervailing considerations, however, suggest that this pattern may have no substantial significance. First, the trial testimony suggested that fighting real fires requires pacing and caution rather than continuing exertion at maximum physical capacity, 536 F. Supp. at 212 & nn. 23-24, and it

may be that the relatively slower times of those older than 29 reflected to some extent their greater experience. Second, as suggested by the district court in expressing its reservations as to whether a test tested by incumbents might not be unfair to the plaintiff class, see note 7 supra, it may be that as a result of the incumbents' firefighting experience, the performance times of all of the test-testers were faster, even without maximum exertion, than would be the times of persons of comparable ages without any experience. In any event, it does not seem to us unreasonable for the City to have sought to determine the average performance of actual job incumbents, and we note that the EEOC Guidelines expressly recommend that a cutoff score "should normally be set so as to be reasonable and consistent with normal expec-

tations of acceptable proficiency within the work force." 29 C.F.R. § 1607.5(H).

We are similarly unmoved by UFA's claim that the interim test does not adequately measure upper body strength. We note, preliminarily, that the evidence suggested, and the court found, that the most important physical attribute for a firefighter is not strength but stamina. 536 F. Supp. at 188, 191 n.6, 204, 212. Moreover, it is plain that the interim test does require upper body strength. It requires, for example, that candidates raise a 58-pound, 20-foot ladder from the ground to a vertical position, lift and stretch an 80-pound hose, and drag a 145-pound dummy for some distance. The 145-pound dummy is 82 pounds heavier than the 63 pounds that the record indicated was the average weight that firefighters were most often

required to drag and, indeed, is 20 pounds heavier than the average extreme weight that firefighters were only occasionally required to drag. See id. at 191. The 80-pound hose is 17 pounds heavier than the weights the record indicated firefighters were most often required to lift. Id. Thus, the interim test required, during a short amount of time (not more than 4:09 minutes), the use of greater upper body strength than is required on the average in fighting real fires.

Finally, we need not dwell on UFA's contention that the interim test is inadequate because it sets lower standards than did Exam 3040. Exam 3040 was found to be discriminatory and not job-related. It cannot rationally be used as a benchmark against which other tests may be judged.

The orders appealed from are affirmed.

Brenda BERKMAN, on behalf of herself and a class consisting of all similarly situated women, Plaintiff,

v.

The CITY OF NEW YORK; Edward Koch, individually and as Mayor of the City of New York; New York City Fire Department; Augustus Beekman, individually and as Fire Commissioner of the City of New York; New York City Department of Personnel; Michael Nadel, individually and as Director of Personnel of the City of New York; Thomas Roche, individually and as former Director of Personnel of the City of New York; Civil Service Commission of the City of New York, Defendants.

No. 79 C 1813.

United States District Court,
E.D. New York

March 4, 1982

Women's Rights Clinic of the Washington Square Legal Services, Inc., by Laura Sager, and Debevoise & Plimpton, by Robert L. King, Bart R. Schwartz, Jeffrey N. Drummond, Kathryn Quirk, New York City, for plaintiff.

Frederick A.O. Schwarz, Corp. Counsel of the City of New York, by Norma Kerlin, Thomas C. Crane, Gary P. Shaffer, Asst. Corp. Counsels, New York City, for defendants.

Norman Teitler, Rego Park, N.Y., attorney for intervenor, Uniformed Firefighters Association.

MEMORANDUM DECISION AND ORDER

SIFTON, District Judge.

This is an action brought pursuant to Title VII of the Civil Rights Act of 1964, as amended, 42 U.S.C. § 2000e et seq.; the Civil Rights Act of 1871, 42 U.S.C. § 1983; the fourteenth amendment to the United States Constitution; and Section 296 of the New York Human Rights Law (Executive Law), seeking declaratory and injunctive relief and damages to redress alleged sex-based discrimination against plaintiff and the class she represents in connection with the physical test portion of New York City's Examination 3040 ("Exam 3040") for the entry level position of firefighter in New York City.

Plaintiff, Brenda Berkman, is a twenty-nine year old woman who passed the written portion of Exam 3040, but failed the physical test portion, which she took on February 22, 1978. The class she represents consists of 410 women who took the written portion of Exam 3040 and are alleged either to have taken the physical portion of Exam 3040 and failed it or to have been deterred from taking it as a result of sex discrimination by defendants. Defendants are the City of New York, its Mayor, the City's Fire Department and its Commissioner, the City's Personnel Department, its Director and former Directors, and the Civil Service Commission of the City of New York. Appearing as Intervenors pursuant to this Court's Order of April 10, 1981, are the Uniform Firefighters Association ("UFA") and the Uniform Fire Officers Association

("UFOA"), both employee organizations representing current job incumbents.

The trial of this case was conducted before the undersigned, sitting without a jury, over several weeks between September and December 1981. For the reascns set forth herein, I conclude that the physical portion of Exam 3040 discriminated against plaintiff and the class she represents on the basis of their sex and that injunctive relief is appropriate prohibiting further use of the eligibility list established pursuant to Exam 3040 except on a showing of compelling necessity, directing the preparation of a new physical exam which does not discriminate against women, and awarding plaintiff interim and other relief to secure compliance with the requirements of the Civil Rights Act of 1964, as amended. What follows sets

forth the findings of fact and conclusions of law on which these determinations are based, as required by Rule 52(a) of the Federal Rules of Civil Procedure.

Jurisdiction over plaintiff's Title VII action exists under 42 U.S.C. § 2000e-5. Plaintiff, by filing her complaint with the Equal Employment Opportunity Commission on May 16, 1978, has complied with the time limitations imposed by that section with respect to all of the named defendants except the defendant Fire and Personnel Departments and the Director of the latter Department, who were not named in plaintiff's administrative complaint. As to those defendants not named in plaintiff's administrative complaint jurisdiction exists since there is substantial identity between those defendants who were named

in the conciliation proceeding and those not named and since those not named had notice of the pendency of the conciliation proceedings. Vulcan Society of Westchester County v. City of White Plains, 82 F.R.D. 379 (S.D.N.Y. 1979).

Jurisdiction exists over plaintiff's claim under 42 U.S.C. § 1983 and the fourteenth amendment by virtue of 28 U.S.C. § 1343. Pendant jurisdiction exists over plaintiff's claim under New York State's Human Rights Law.

Background

As of the end of 1980, New York City employed in excess of 285,000 persons. Of these, approximately 175,000 (including those in the City's Fire Department) were hired by the Department of Personnel. The balance were employed by independent agencies such as the Board of

Education and the Off-track Betting Corporation.

Of the approximately 175,000 employed by the Department of Personnel,- approximately 168,000 (including those in the City's Fire Department) were in the competitive class, a class including all positions for which it is practicable to determine the merit and fitness of candidates by competitive examination.

Employment statistics for the uniformed force of the New York City Fire Department for the period 1973 through 1981 were as follows:

Year	Firemen	Officers	Total
1973	10,720	2,548	13,394
1974	10,426	2,538	13,091
1975	9,089	2,347	11,548
1976	8,304	2,267	10,662
1977	8,847	2,329	11,271
1978	8,513	2,375	10,979
1979	8,966	2,404	11,466
1980	8,765	2,460	11,048
1981	9,042	2,563	11,616

The Fire Department of the City of New York is charged with responsibil-

ty for extinguishment, prevention and
nvestigation of fires occurring in the
ity. The most recent available statis-
ics for operational firefighting inci-
ents (1979) indicate a total of close to
60,000 incidents annually requiring Fire
epartment attention. The largest number
these incidents have been false alarms
62,529). A total of 43,072 of these
cidents involved structural fires, of
ich the largest number occurred in
sidential buildings (31,504), the next
rgest, in vacant buildings (5,698),
llowed by fires in commercial (4,227)
d public structures (1,643). Non-
ructural fires (71,298) exceeded struc-
ral fires. In addition to its ordinary
refighting duties, the Fire Department
s, in 1979, called upon to respond to
ose to 72,250 other emergencies of
rious types.

The Fire Commissioner is the head of the Department, responsible for policy decisions. The Chief of Department is charged with operational responsibility. Under the Chief of Department the uniformed force of the Department consists of the assistant chiefs, deputy assistant chiefs, deputy chiefs, battalion chiefs, chief medical officer, medical officers, chaplains, captains, lieutenants, marine engineers, pilots, and firemen.

The uniformed force is organized into divisions, battalions, companies, and other operational units. A division is composed of one or more battalions and is commanded by a battalion chief. A company is composed of its captain, one or more lieutenants, and the firefighters assigned to it.

The command structure of Department operations consists of borough commands in each of the five boroughs of New York City. Each borough command (with the exception of Staten Island) is composed of several divisions, the number varying in each borough. The borough commanders report to the Chief of Operations.

Companies are divided into engine companies, ladder companies, rescue companies, and marine companies. An engine company consists of a pumper apparatus and an assigned complement of personnel (typically, one captain, three lieutenants, and 20 or 25 firefighters). One platoon (typically, one officer and four or five firefighters) is assigned to the apparatus on each tour of duty. The pumper apparatus consists of pumping equipment and carries lengths of

hoseline, nozzles, and other equipment. The engine company is primarily responsible for the extinguishment of fires.

A ladder company consists of a ladder truck apparatus and an assigned complement of personnel (typically, one captain, three lieutenants, and 25 firefighters). One platoon (typically, one officer and five firefighters) is assigned to the apparatus on each tour of duty. The ladder truck apparatus includes an extension ladder, portable ladders, tools and equipment for entry and ventilation of burning buildings, and a variety of other equipment. The members of the ladder company perform operations at a fire scene exclusive of extinguishment, including forcible entry, ventilation, search and rescue, and overhauling, the process of seeking hidden

sources of fire after the visible fire is extinguished.

There are at present 208 engine companies and 138 ladder companies in the Department. These are currently housed in 233 firehouses, most of which quarter one ladder company and one engine company.

In addition to engine companies and ladder companies, there are rescue and marine companies. A rescue company consists of the rescue apparatus and an assigned complement of personnel (one captain, three lieutenants and 25 firefighters). A platoon (one officer and five firefighters) is assigned to the apparatus on each tour of duty. The members of a rescue company are assigned on an ad hoc basis at fires to augment ladder or engine company personnel. They also perform specialized search and

extrication tasks, utilizing special tools carried on the apparatus for use in emergencies. There are four rescue companies in the City, one assigned to each borough, except Staten Island. Rescue companies respond to multiple alarm fires and unusual emergencies.

A marine company is responsible for extinguishment of fires on the waterfront and in New York Harbor. A marine company consists of a captain, three lieutenants, four pilots, eight marine engineers, one wiper, and either ten to twenty firefighters. Forty-five firefighters are assigned to the marine division.

Under Section 487a-3.0a. of the New York City Administrative Code (the "Administrative Code"), a member of the Fire Department must be a citizen of the United States, must be able to read and

write the English language with understanding, and must have passed his or her eighteenth, but not twenty-ninth, birthday on the date of the filing of an application for civil service examination. No person will be appointed unless twenty-one years of age at the time of appointment. There is at present no restriction on the sex of applicants.

There are four grades of firefighter. Section 487a-7.0 of the Administrative Code provides in this connection:

> "Members of the uniformed force, upon appointment shall be assigned to the fourth grade; after one year of service in the fourth grade, they shall be advanced to the third grade; after one year in the third grade, they shall be advanced to the second grade; after one year in service in the second grade, they shall be advanced to the first grade; and they shall in each instance receive the annual pay or compensation of the grade to which they belong."

The basic starting salary for firefighter fourth grade is currently $17,949.

Under Section 487a-4.0 of the Administrative Code, preliminary to a permanent appointment as a firefighter, there is a period of probation of one year. Probationary firefighters undergo a six-week training program in the fundamentals of firefighting at the beginning of their probationary period.

The Department of Personnel of New York City has responsibility for the hiring of Fire Department personnel. When staffing needs arise for the position of entry-level firefighter, the Fire Department advises the Department of Personnel and requests the scheduling of an entrance examination. The examination is prepared by the Department of Personnel with the assistance of the Fire Department. Scheduling, notification,

receipt of applications, and administration of the examination are the responsibility of the Department of Personnel. The examination produces an eligible list of individuals who have passed the examination. Thereafter, the Department of Personnel and the Fire Department conduct further investigations, including background investigation, medical examination, and interviews, of sufficient members of the eligible list to satisfy projected hiring requirements. Candidates who satisfactorily complete the post-examination investigations are deemed qualified. In order to fill vacancies, the Fire Department will request certification of qualified eligible candidates from the Department of Personnel. Upon issuance of such certification, the Fire Department appoints candidates from the qualified eligible list.

Examinations for the Position of Fireman Prior to Exam 3040

From 1960 to 1971 the Department of Personnel administered five examinations for the position of fireman. None of these was open to women. These were the five examinations which immediately preceded Exam 3040, which is at issue in this lawsuit.

From October 5, 1960, to January 31, 1961, pursuant to Notice of Examination 9010, the Department of Personnel accepted applications for the position of fireman. A written test for 9010 was administered March 3, 1961, and the eligible list was promulgated January 1, 1962. No physical test appears to have been administered.

From July 5, 1962, to October 2, 1962, pursuant to Notice of Examination 9606, the Department of Personnel accepted applications for the position of

fireman. A written test for 9606 was administered December 1, 1962, followed by a qualifying physical test. The physical test was graded on a pass/fail basis; the written test had a passing score of 70%. The eligible list resulting from this process was promulgated June 26, 1963.

From December 3, 1963, to January 27, 1964, pursuant to Notice of Examination 9984, the Department of Personnel accepted applications for the position of fireman. A written test for 9984 was administered on May 19, 1964, followed by a competitive physical test from which an eligible list was promulgated. The written examination and the competitive physical examination were given equal weight. An average of 70% was required on the results of both writ-

ten and physical tests in order to qualify for appointment.

From April 3, 1968, to May 31, 1968, pursuant to Notice of Examination 7060, the Department of Personnel again accepted applications for the position of fireman. The competitive written test for 7060 was administered June 15, 1968, followed by a qualifying (pass/fail) physical test from which an eligible list was promulgated August 20, 1968. A 70% score on the written test was required to pass.

From February 3, 1971, to August 18, 1971, pursuant to Notice of Examination 0159, the Department of Personnel accepted applications for the position of fireman. The written test for 0159 was administered on September 18, 1971, followed by a qualifying (pass/fail) physical test, and an eligi-

ble list was promulgated January 18, 1973. Again, a 70% score on the written test was required to pass.

The Vulcan Society Litigation

On January 12, 1973, immediately preceding the promulgation of the eligible list resulting from Exam 0159, the Vulcan and Hispanic Societies, representing black and Hispanic firefighters, commenced an action in the United States District Court for the Southern District of New York pursuant to 42 U.S.C. §§ 1981 and 1983 and sought a preliminary injunction to restrain the New York City Personnel Department, Civil Service Commission, and Fire Department from appointing firemen from List No. 0159. Vulcan Society v. Civil Service Commission, 73 Civ. 201 (Weinfeld, J.).

Thereafter, on June 12, 1973, Judge Weinfeld issued an opinion declar-

ing Examination 0159 unconstitutional under the equal protection clause because of its discriminatory impact on blacks and Hispanics and enjoining defendants from making further appointments based on the examination's results, except upon a showing of compelling necessity. 360 F. Supp. 1265, aff'd in part and remanded, 490 F.2d 387 (2d Cir. 1973).

Subsequently, on August 2, 1973, Judge Weinfeld issued an order directing defendants, inter alia, to proceed expeditiously to prepare a new examination for the position of fireman which did not discriminate against blacks and Hispanics in accordance with professionally accepted methods of test preparation.

Retention of AIR

The City had previously placed itself in a position from which it was

able to commence compliance with Judge Weinfeld's direction with some rapidity. In 1970 Congress had enacted Public Law 91-648, known as the Intergovernmental Personnel Act ("IPA"), 84 Stat. 1909 (1970), which described itself as follows:

> "An Act to reinforce the federal system by strengthening the personnel resources of State and local governments, to improve intergovernmental cooperation in the administration of grant-in-aid programs, to provide grants for improvement of State and local personnel administration, to authorize Federal assistance in training State and local employees, to provide grants to State and local governments for training of their employees, to authorize interstate compacts for personnel and training activities, to facilitate the temporary assignment of personnel between the Federal Government, the State and local governments, and for other purposes."

Pursuant to Title II of the IPA, "Strengthening State and Local Personnel Administration," the United States

Civil Service Commission was authorized to make grants to state and local governments to enable them to strengthen their staffs by improving personnel administration. Pub. L. No. 91-648, § 201, 84 Stat. 1911 (1970).

In January 1973, the City's Personnel Department had determined to take advantage of this federal program by applying for and receiving IPA funding for development of personnel tests and personnel test validation procedures. With the promise of such funding, the City, in January 1973, entered into two contracts with a Washington, D.C. consulting firm named American Institutes for Research ("AIR") -- one directed towards developing a method of constructing valid physical selection tests for physically demanding city jobs, the other

directed towards the development of suitable written examinations.

Since IPA grants are awarded to support broad research projects and not to pay for the development of an operational selection procedure to select employees for a particular job, the 1973 City contracts looked to the development of generalizable approaches to job analysis, test development, and validation which could be applied to a wide variety of city jobs. In accordance with IPA requirements, both contracts were for one year only. However, each contract was thereafter renewed annually for the following two years.

Although general in their overall objectives, both contracts provided that AIR would study three job titles in particular in order to develop more broadly applicable methods. In March

1973 the decision was made that the three jobs to be studied would be fireman, sanitationman, and parking enforcement agent.

Specifically, as of January 1, 1973, the City contracted with AIR for the "Development of Physical Test Validation Procedures" to be applied to jobs with physical demands. The contract was authorized under IPA Grant No. 72-14. The full amount of the contract ($37,476) was to be paid with funds received from the U.S. Civil Service Commission. The term of the contract was January 1, 1973, through December 31, 1973, later extended to March 31, 1974. The second year of the City's contract with AIR for the "Development of Physical Test Validation Procedures" was carried out under IPA Grant No. 73-NY-27c. The full amount of the contract ($22,500) was to be payable

with funds received from the U.S. Civil Service Commission. The term of the contract was April 1974, through December 31, 1974, later extended to February 15, 1975.

Also, as of January 1, 1973, the City entered into a contract with AIR to develop written test procedures. Programs were to be developed which, with a minimum number of modifications, might be utilized for most entry-level City positions and some promotional positions. The contract was authorized under IPA Grant No. 72-5, and the full amount of the contract ($59,992) was to be payable with funds received from the U.S. Civil Service Commission. The term of the contract was January 1, 1973, to December 31, 1973, but was extended to March 31, 1974. The second year of the City's contract with AIR for the develop-

ment of written test validation procedures was carried out under IPA Grant No. 73-NY-28c. The full amount of the contract ($19,956) was to be payable with funds received from the U.S. Civil Service Commission. The term of the contract was April 1, 1974, through December 31, 1974, later extended to February 15, 1975.

While the parties to the contracts had contemplated that the overall term of the projects undertaken by AIR would last three years and while IPA grants (No. 75-NY-04c and No. 75-NY-05c in the amounts of $19,922 and $15,000 for the physical and written projects, respectively) to be paid in full by the U.S. Civil Service Commission had been issued, both contracts were cancelled by the City, effective September 30, 1976. These cancellations were explained by

witnesses for the City at trial as attributable to the City's fiscal crisis.[1] Whatever the merits of this explanation the cancellations, as will be seen, had the unfortunate consequence of depriving the City of an opportunity, at least as part of AIR's IPA-funded work,

---

[1] What the fiscal crisis prevented was the hiring of firemen, parking enforcement agents, and sanitationmen who were to be the subjects of the validation study. Since, as discussed infra, the AIR study contemplated a criterion-based validation study to find out whether the test scores accurately predicted actual job performance, the failure by the City to do any hiring is said to have made such a criterion study impossible. The fiscal crisis does not, however, entirely explain why a concurrent validation study not involving new hiring, as described in an AIR letter to the City of September 22, 1976 (Plaintiff's Exhibit [P.X.] 141), could not have been performed with IPA money. What this letter also demonstrates is that no validation of the physical portion of the exam of any sort had been done prior to the time the letter was composed.

to produce a criterion-based validation study of the physical test for fireman which is at issue in this litigation -- that is, a follow-up study of those who passed the physical test to determine how well they performed as firefighters in order to decide if the test accurately predicted job performance.

In all events, in the period immediately following the June 1973 decision in the Vulcan case, with AIR already performing pursuant to these two contracts and, coincidentally, investigating the position of City firefighter as one of the three job titles from which a generalizable approach was to be developed, the City understandably turned to AIR for the specific purpose of developing a new entrance level examination for firefighters as required by Judge Weinfeld's decision.

Accordingly, on October 25, 1973, the City entered into a contract with AIR for the purpose of preparing an examination for firemen in the New York City Fire Department. The contract was not to exceed $32,500, and the work was to be completed by April 12, 1974, although this date too was later extended. In the contract AIR agreed, inter alia, to comply in the course of its work with EEOC Guidelines on employee selection pertaining to fair employment, to conduct a job analysis of the position of fireman, using material and expertise it had acquired on its two other contracts with the City, and to prepare, based on the job analysis, a notice of examination, test plan, written test, medical standards, and minimum standards and procedures to be followed in administration of the physical test. Precisely what por-

tions of its earlier work in connection with the 1972 contracts were, in fact, drawn upon by AIR in developing Exam 3040 pursuant to this contract is the subject of considerable debate between the parties. However, since all parties are agreed that at least the physical abilities analysis ("PAA") conducted by AIR pursuant to the 1972 physical test contract formed a central part of the job analysis eventually relied upon to develop the firefighters' exam, it is appropriate to discuss that aspect of the job analysis first.[2]

---

[2] In addition to the PAA, AIR analyzed all three jobs, that is, fireman, sanitationman, and parking enforcement agent, for purposes of both the written and physical exam projects by means of the following methods: job inventory, critical incidents analysis, position analysis questionnaire, and physical demands analysis. AIR personnel also met and conferred with Fire Department officials and observed the probationary and training schools
(Footnote continued)

(Footnote 2 continued from previous page) and firefighting operations on a number of more informal occasions.

Since defendants now seek to rely on selections from some or all of these job analysis techniques, other than the PAA, to validate Exam 3040 and since there is in each of them some reference to the kind of observable behavior and ratings of importance which are missing from the PAA, it is important to understand what role these other analysis techniques actually played in the preparation of the physical test portion of Exam 3040. In general, as stated by AIR in its Final Report:

> "A variety of job analysis procedures were reviewed in order to chose [sic] techniques appropriate for the present effort. Many of the standardized techniques such as the Job Inventory, Abilities Analysis, Functional Job Analysis, and the Position Analysis Questionnaire often include cognitive, perceptual, and psychomotor dimensions in addition to physical proficiency ones and are designed to assess requirements for total job performance. Since the focus of this project involves the identification of physical job requirements exclu-
>
> (Footnote continued)

contracts specified that AIR's work for the City would build on previous research

---

(Footnote 2 continued from previous page)

> sively, total job analyses were evaluated in terms of how detailed and comprehensive the physical abilities portion of these instruments were [sic]. The physical portion of Abilities Analysis represents one detailed categorization of the physical performance domain empirically derived from extensive factor analytic and correlational analyses (see Fleishman, 1964). One other technique reviewed, called Physical Demands Analysis, concentrates particularly on physical attributes of job performance.
>
> "Rather than pre-judge the relative effectiveness of describing physical requirements of jobs by either physical abilities or physical demands both techniques were utilized and later evaluated in terms of the results which they produced."

Technical Report No. 3 (P.X. 87), prepared by AIR, confirms that the job analysis methodology ultimately used with respect to the physical requirements of the fireman's job involved "the utilization of two job assessment techniques": the PAA and the PDA. Between these two job analyses tech-

(Footnote continued)

and methodology developed by Dr. Edwin Fleishman, an industrial psychologist,

---

(Footnote 2 continued from previous page)
niques, the PAA was ultimately selected for the reason that, relying on Fleishman's work, it already had associated with each ability identified an appropriate test thus, theoretically, facilitating the task of test preparation. According to AIR: "Data derived from Physical Demands Analysis help to determine the exact nature of the tests to be used which are diagnostic of the abilities selected."

Despite this clear evidence that AIR itself intended to rely exclusively on the PAA and the PDA to construct the physical portion of Exam 3040, defendants sought at trial to use evidence of observable behaviors reported as part of at least three other job analysis techniques to demonstrate the content validity of the physical exam, namely, the job inventory, the critical incidents analysis, and the position analysis questionnaire. Accordingly, it is appropriate to say a word about AIR's use of these three techniques.

The job inventory involved by far the most far reaching inquiry made by AIR of all of its job analysis efforts. The effort started with the creation of a checklist of 269 work activities involving both physical and mental abilities derived from training man-
(Footnote continued)

then the head of AIR's principal office. Dr. Fleishman's previous research in

---

(Footnote 2 continued from previous page) uals and meetings with Fire Department officials. These were then divided into eight categories, e.g., public education and community relations, inspection, investigation and enforcement, miscellaneous firehouse activities, firefighting, and other emergency and support operations. Each activity and each category of activity was then rated in terms of its difficulty, the time spent on it, and its importance, first by 20 job incumbents, consisting of ten firemen and ten officers, and then by some 600 firemen who completed the job inventory questionnaire. The result achieved from this inventory was the general conclusion that "firefighting and other emergency and support operations" rated first in terms of job importance, difficulty, and time consumed. The ratings given for the top quartile of the 269 job activities were found to confirm this general conclusion. Unfortunately, the 90 job activities rated most difficult and most important (e.g., in order of importance, "drive apparatus to and from scene, operate apparatus at scene, select appropriate tool or piece of equipment, make and unmake couplings and connections, estimate lengths of hose needed, stretch hose vertically. operate low pressure hydrant," etc., etc.) were not analyzed (Footnote continued)

physical fitness measurement is reported in his book, The Structure and Measure-

---

(Footnote 2 continued from previous page) by AIR either at the time of the test's preparation or at trial in terms of physical abilities involved in the activity or in terms of the tests administered as part of Exam 3040. In place of such an analysis, defendants have simply asserted on the basis of quite superficial resemblances between a few of the many job activities listed in the job inventory and certain of the tests actually administered as part of Exam 3040 that the tests were content validated by the job inventory (e.g., the job activity described as "perform search operations" is said to be "connected" to the window-ladder-window component of the agility test, Defendants' Interim Post Trial Brief at p. 29). In sum, defendants appear to be engaged in ex post facto rationalization of tests selected for quite different purposes, drawing on aspects of AIR's job analysis which were never intended to be put to such use.

The so-called critical incidents analysis involved representatives of AIR sitting down with the two groups of firemen and officers, the same groups as those used in the job inventory, "to discuss examples of very effective and ineffective job performance and attributes believed to be important for job performance." This process

(Footnote continued)

ment of Physical Fitness (1964). The book describes a method of factor analy-

(Footnote 2 continued from previous page) resulted in two lists describing in narrative form "examples provided of very effective and ineffective performance by individual firemen," and "attributes believed to be important for the job of fireman." Among the 32 "attributes" listed as "important" (along with "need initiative and imagination," "need to administer first aid," "need to make decisions and follow through," and "know the use of tools") are "strength" (including "dynamic strength" and "explosive strength"), "stamina," "balance," "coordination," "extent flexibility," "dynamic flexibility," and "speed of limb movement." Under each of these general categories of "attributes" (which appear to have been drawn directly from the language of the PAA rather than from a job study), a few examples are given, e.g., "an example of explosive strength is when the man must raise a ladder." However, here, as with the "comments" section of the PAA, there is a lack of any systematic effort to determine whether the respondents understood the abilities being discussed in the same sense in which they were used in the PAA. Moreover, there is a failure to assess the relative importance of the specific job behaviors reported vis a vis the overall job and vis a vis each other.

(Footnote continued)

sis by which Dr. Fleishman identified eight basic physical fitness "factors" as comprising the domain of human physical performance. Dr. Fleishman isolated these factors by applying a factor-analysis methodology described in the book to test scores on various physical

---

(Footnote 2 continued from previous page)

The results of the Position Analysis Questionnaire ("PAQ"), which are at a level of generality equal to that of the PAA, in fact raise serious questions about the reliability of the PAA results. The ordering of importance of the physical attributes of stamina, explosive strength, dynamic strength, and static strength based on the results of the PAQ appears strikingly different from the ordering accorded those attributes based on the results of the PAA (D.X. GG p. 31). Moreover, in those few instances in which defendants claim that the results of the PAQ confirm the results of the PAA, it is to be noted that the definitions of the matters being inquired about differ from one analysis instrument to the other and are, in fact, measured on a different scale. (Cf. D.X. BB, p. 28 with D.X. GG, p. 27.)

fitness tests -- all, coincidentally, administered to male subjects.

Factor analysis is a statistical technique for determining the minimum number of factors necessary to account for the intercorrelations among a set of variables. Through factor analysis, Dr. Fleishman isolated clusters of physical tests that appeared to measure the same underlying ability. Each cluster represents a factor. The correlation between a test and a factor is called factor loading.

The eight factors identified by Dr. Fleishman in his 1964 book are static strength, explosive strength, gross body equilibrium, extent flexibility, gross body coordination, dynamic strength, dynamic flexibility, and speed of limb movement. A ninth physical fitness factor, stamina, was subseqently adopted by

Dr. Fleishman as an additional physical ability from tests reported by other researchers in the field of industrial psychology.

Dr. Fleishman's 1964 book also identifies the factor loadings of each of the physical fitness tests on each factor and contains recommendations as to which tests best measure each hypothesized underlying ability. Tests were recommended on the basis of being "factorially pure," i.e., loading highest on the factor tested and not loading significantly on other factors. The book also contains data on the reliability of the various tests recommended.

In 1970 and 1971 Dr. Fleishman undertook further research to identify the full range of all human abilities. Based on this work, he reported that there are 37 human abilities: 14 cogni-

tive abilities, nine psychomotor abilities, five perceptual abilities, and the nine physical abilities enumerated above. Having identified the 37 human abilities, Dr. Fleishman and his colleagues developed a system of rating scales, known as the abilities analysis, which could be used by raters to indicate on a scale of one to seven how much of a given ability is needed to perform a particular task.

The rating scales include references to particular concrete instances of human behavior thought to exemplify the ability under consideration and a particular degree of that ability. These concrete examples were intended to serve as "anchors" at intervals along the scale to assist the rater in using the instrument. These behavior anchors were developed based on research with groups of psychologists who devised the task an-

chors in group sessions and later participated in trying out the scales by using them to rate various tasks. The anchors are descriptions of various kinds of human activity and do not reflect tasks commonly performed by any particular occupational group.

For example, anchors under the physical ability of stamina include, at the top of the seven-point scale, "swim the English Channel"; at a point mid-way between five and six on the scale of seven, "climb the stairs to the top of the Statue of Liberty"; and at the bottom of the scale, "walk to the corner grocery." Under gross body coordination the anchors are "do a skilled ballet dance like Swan Lake"; "a runner jumps a series of ten three-foot hurdles"; and "make a lay up shot in the basketball game." Under extent flexibility the anchors are

"side show lady bends herself into a pretzel shape and other unusual positions"; "do a modern dance"; and, at the bottom of the scale, "reach for the control lever of a drill press." Using these scales the person completing the abilities analysis is asked to state "how much" of each ability is needed to do the task under study.

In addition to the task anchors, the test instrument gives a definition of the ability, e.g.:

> "Static Strength. This ability involves the amount of muscle force (power) used against an object to lift, push or pull the object. Force is used without stopping up to the amount needed to move the object. This ability can involve different muscle groups like the hand, arm, back, shoulder, and leg. This ability does not involve the use of force over a long time. This ability does not deal with the number of times the act is repeated." (Emphasis in original)

Further, an effort is made in each instance to differentiate the ability under consideration from other abilities with which it might be confused. Thus, static strength is contrasted with explosive and dynamic strength and with stamina as follows:

> "HOW STATIC STRENGTH IS DIFFERENT FROM OTHER ABILITIES: Use force without stopping to lift, push or pull objects
>
> vs.
>
> Explosive Strength (2): Gather energy to move the body's own weight or objects with short bursts of force.
>
> Use force to lift, push or pull objects
>
> vs.
>
> Dynamic Strength (3): Use muscle power (force) repeatedly or without stopping to hold up or move the body's own weight. Does not involve the use of force over a long time
>
> vs.
>
> Stamina (4): Do physical work over a long time; involves heart and

blood vessels resisting becoming tired."

Further, the persons performing the abilities analysis are asked to complete a sheet entitled "comments" on which they are to state separately with respect to "usual tasks" (described as those performed every day or almost every day on the job) and with respect to "special tasks" -- "what the worker does" and "under what circumstances the worker does it." It is principally based on these comments that defendants contend that the PAA was content validated as grounded in the important observable behaviors of firefighters.

The Job Analysis of the Physical Requirements of the Firefighter

In November 1973, AIR administered the PAA to 23 firemen and 12 fire officers. This use of the PAA was the

first time that the instrument had ever been used in an actual analysis of an entire job as opposed to the analysis of a specific task. The firemen and officers were asked to rate the amount of each ability required by all of the usual tasks of their job and by all of the special tasks of their job on a scale of 0 to 7. Usual tasks were, as noted above, defined as tasks performed every day or every other day. Special tasks were defined as tasks performed less often. The PAA questionnaire also asked the raters to provide written comments, giving examples of both usual and special tasks which required each ability. These comments provide some idea of the various specific tasks which the raters were considering in answering the questions directed to the rather abstract abilities they were asked to rate.

For some reason, most probably because of the difficulties presented by the complex analysis instrument, relatively few of the 35 firemen who did the PAA were able to think of special tasks -- that is, ones performed on a less frequent basis than every day or almost every day on the job. As a result, noting that the rank orderings for special tasks would be influenced by very small differences in mean ratings and calling them "secondary ratings," AIR determined to rely on the ratings given for every day tasks and noted that the results for the special tasks analysis "should be viewed with caution and considered merely suggestive."[3]

---

[3] As stated in AIR's final report: "An examination of the ratings of abilities for special tasks and narrative descriptions revealed that these data were entirely too fragmentary and divergent to be of any value here."

Using the responses given for usual tasks, the abilities were then ranked as follows, based on the mean ratings for both officers and firemen:

Ability	Mean Score on Scale of 7
Stamina	5.41
Static Strength	4.64
Explosive Strength	4.41
Gross Body Equilibrium	4.27
Extent Flexibility	4.26
Gross Body Coordination	4.04
Dynamic Strength	4.03
Dynamic Flexibility	3.74
Speed of Limb Movement	3.41[4]

---

[4] Given the bounded nature of the scale and the small size of the sample used, it is questionable whether the differences between the ratings reach the level of statistical significance. This becomes even more apparent when one considers the results of ratings for the abilities used in special tasks:

Ability	Mean Score on Scale of 7
Stamina	5.86
Extent Flexibility	5.82
Dynamic Strength	5.81
Explosive Strength	5.67
Gross Body Equilibrium	5.64
Gross Body Coordination	5.54
Static Strength	5.50

(Footnote continued)

The AIR effort was not only the first occasion on which the entire job was analyzed using the PAA; it was also the first time job incumbents, as opposed to professional scientists, were asked to perform the rather abstract abilities analysis. In all earlier uses of the test instruments to rate specific tasks, the ratings had been made by trained psychologists, familiar with the mechanisms of the analysis, who had at least some acquaintance with the definitions they were being asked to apply and some of whom had, in fact, participated in the

---

(Footnote 4 continued from previous page)

Dynamic Flexibility	5.43
Speed of Limb Movement	5.29

AIR itself noted this deficiency in its final report: "It should be noted . . . that the differences between mean ratings for adjacently ranked abilities are very small whether considered within or across groups of ranked abilities."

preparation of the test anchors used in an effort to make the analysis concrete. Given the complexity of the rating scales and the nature of the task anchors, the analysis is clearly capable of producing a different result when used by different individuals with different backgrounds. Thus, the significance of an anchor for static strength (which is set at a level of 5.8 on a scale of 7) described as loading five full fifty-gallon drums into a truck may be quite different for a person who has done the task described than for a person who has not. In fact, there is considerable evidence that actual confusion and misuse of the instrument did occur.

As already noted, a large number of respondents failed to find the selected abilities present at all in special firefighting tasks, i.e., those

occurring less frequently than once every one or two days. Moreover, many of the same tasks were considered usual tasks by some respondents and special tasks by others. Thus, "overhauling using ladders," "ventilation," "lifting people to stretchers," "moving large refrigerator or furniture in fire building," "making rescue of person or persons of varying weights under adverse conditions," and "forcible entry" are all listed as special tasks by some respondents. Other respondents describe essentially the same tasks, e.g., "moving usual household items such as refrigerator in a time of heavy fire conditions," "overhauling at a fire," "needed when forcing entry through doors or walls under fire conditions," "moving heavy refrigerators, furniture, unconscious persons," "to remove person from smoke-filled room or apartment,"

"force doors, vent," as a part of usual operations.

Moreover, many of the comments list the same activity or job behavior as an example of all or almost all of the abilities inquired about in the PAA questionnaire. For example, "functions at fire operations" and "stretching a hose" is given as an example of each of the given abilities in one fireman's comments. While this may reflect a correct understanding of the abilities being asked about (it is possible, for example, given the complexity of body movements involved in stretching a hose, to find at one point or another in the evolution an example of static strength, explosive strength, dynamic strength, stamina, extent flexibility, and gross body coordination), it is difficult to say, given the complexity of the example of

job behavior offered in the comment, whether the example reflects a correct understanding of the ability or not. Nor is it clear that one has given concrete support for the degree of importance attributed to the ability when one has cited a complex job behavior involving to some unspecified degree the ability inquired about, but also involving other abilities.

This point may be made clearer by examining the definition of dynamic strength given in the PAA, along with the contrasting definitions provided, and then considering a number of examples given for that ability in the comments of firemen.

Dynamic strength is defined in the PAA instrument as follows:

> "This ability involves the power of arm and trunk muscles to hold up or move the body's own weight repeatedly or at one time without stopping. This ability involves muscles resisting getting tired when having to repeatedly or without stopping to [sic] hold up or move the body's own weight." (Emphasis in original)

Contrasts are then provided between dynamic strength and static strength (involving the use of force to lift, push or pull objects), explosive strength (involving short bursts of muscle force to move the body's own weight or objects), and stamina (defined as work over a long time, involving "heart and blood vessels resisting becoming tired").

The following comments by firemen, listing job behaviors thought to exemplify dynamic strength, are taken in order from a list prepared by AIR from the firemen's comments:

"Pull hose up six flights in smoke, pull ceilings; overhauling operations; carry persons up or down stairs; to and from places.

"Advancing a line into a building, in heat and smoke. Rescue operations. [Underlined comments represent a special task.]

"Advancing a line when charged up two or three floors or when pulling a charged line in a window off the fire escape four floors up.

"Firefighting; climbing stairs with tools and Air Paks on backs; climbing ladders; leaning over roofs to break windows.

"Under heavy fire conditions in rescues, this is pertinent.

"Pulling and running at a fire.

"Act of fighting a fire -- carrying in line, movement from place to place with line while extinguishing fire.

"Needed when advancing charged line at fire. Needed for opening ceilings and walls to allow access to fire to engine company.

"Pulling your own body up to the top of tall bulkheads. Pulling yourself over fences and walls to reach a certain objective (rear office building, adjoining roof tops). Pulling yourself up outside of fire escape to pass fire."

What is noteworthy about this list, when it is compared with the definition for dynamic strength given above, is that none of the groups of examples, except the last, appears to involve the defined ability as anything like its most noticeable characteristic.[5] Indeed, each of the first eight activities described appear to involve first and foremost the contrasted ability of static strength in which force is used to lift, push or pull objects. The contrasted ability of stamina and other unmentioned abilities such as extent flexibility appear as much

---

[5] It should be noted that not all of the abilities inquired about appear, from the examples given in the comments, to have generated the same degree of confusion. The examples given for activities requiring stamina in particular appear to be in accord with the definition given in the PAA, which is in turn a definition which accords well with our common, everyday understanding of the term.

involved in the examples listed as the defined ability of dynamic strength. When these examples are placed alongside others in which the defined ability is almost certainly misunderstood, e.g., "moving Multi-Versal Nozzle from apparatus" as an example of explosive strength, it appears that the PAA and its ranking of abilities are not so well grounded in the observable behaviors of firefighters that one can predict with any confidence from the degree of which the abilities are possessed the capacity of an individual to perform well as a firefighter.

Moreover, despite the small size of the group used to perform the analysis and the failure to select that group on a random basis, no effort appears to have been made -- either on a subjective basis, by weeding out re-

sponses that evinced misunderstanding or inappropriate use of the scales, or on the more objective and systematic basis -- to compute inter-rater correlation coefficients. Inter-rater correlation coefficients are statistical expressions of the degree to which raters agree with each other. Without inter-rater correlation coefficients, a researcher can have only a rough idea with regard to the consistency with which raters ranked the rated abilities. The only inter-rater correlations which do appear to have been calculated -- those comparing the rank ordering by firemen with the rank ordering by officers -- supply further troublesome data, since there appears to have been a significantly greater degree of agreement between the officers and men in the more concrete physical demands analy-

sis ("PDA") than that found with respect to the PAA.

Finally, AIR failed to reconcile the results of the PAA with data from other job analysis techniques. In this connection, it appears appropriate to consider, in particular, the PDA in some detail in part because of the contrast between the PDA and the PAA in the degree to which the analysis is concretely grounded in observable work behaviors.

The Physical Demands Analysis

The PDA is a job analysis method, designed by the U.S. Department of Labor, which describes systematically the physical demands of a job in terms of 11 different activities: strength, climbing, balancing, stooping, kneeling, crouching, crawling, reaching, handling, fingering, and feeling. On November 13,

1973, a panel of 19 firemen and 12 officers under the supervision of AIR used the PDA instrument to describe the job of fireman. The panel was first asked to divide the job into six categories -- sitting, standing, crawling, walking, running, and other -- and to state the percentage of time spent in each activity. The results showed that 71.7% of the firefighter's time was spent in the more physically taxing work positions of standing, walking, running, and crawling.[6]

Next, the panel was asked to analyze the strength demands of the job by stating in terms of "sometimes," "often," and "very often" the need to lift,

---

[6] This high rating for standing, walking, running, and crawling was properly considered by AIR to confirm the high mean rating for stamina given in the PAA.

carry, push or pull five categories of weights from the very light (0-10 lbs.) to the very heavy (100+ lbs.). The results of this exercise showed that firefighters sometimes have to lift weights on the average of 110 lbs., carry weights on the same average, push weights of 161 lbs., and pull weights of 125 lbs. At the other end of the scale, firemen "very often" lift, carry, push or pull weights averaging 63 lbs. Each of these activities -- lifting, carrying, pushing, and pulling -- were then particularized by questions as to what was required to be moved, the circumstances, the tools to be used, and the nature of any special job or task requiring the ability. In addition, respondents were asked to rate the importance of strength in general on a scale of no importance, important, and high importance. Although the answers to

these questions would appear to have been particularly valuable in establishing a content valid test for strength, little use appears to have been made of them in fact, since the PAA provided another system of rating strength as a required ability in terms of three separate components, static strength, explosive strength, and dynamic strength, and since it was the PAA which, as noted below, was principally relied upon.

After dealing with strength in this fashion, the PDA further investigated each of the remaining activities (e.g., climbing, crawling, balancing, fingering, etc.) by asking whether it is present in the job, its importance in terms of the same importance scale referred to above and what climbing, balancing, etc. the worker did (and, in the case of climbing, how far the firefighter

had to climb), the circumstances under which the activity was performed, and the name of any special task requiring the activity.

While AIR made some effort to compare the results of the PDA with the results of the PAA, some of the comparisons smack of an effort at rationalization of apparent conflicts in the results of the two analyses to support use of the PAA. Thus, although climbing and crawling were rated first and fourth out of the eleven behaviors on the PDA, the ability of dynamic strength with which AIR associated them is given a rank of seven out of the nine abilities on the PAA. The differences are to be reconciled, according to AIR, because dynamic strength was ranked third with respect to special tasks on the PAA. Elsewhere, however, as noted above, AIR had cau-

tioned that the general lack of response to the special tasks portion of the PAA made it not a very useful part of the job analysis. As a result, the contrast between the ratings of the PDA and the PAA for dynamic strength remains troubling.

So too, AIR noted that reaching is ranked considerably lower on the PDA than the ability with which AIR associated it on the PAA (extent flexibility). AIR explained this difference as "most likely" reflecting the difference of definition given for the two physical activities. In other areas no comparison at all is made between the ratings achieved on the PAA and on the PDA because the differences of definition between the physical demands involved in the PDA and the physical abilities involved in the PAA make clear that differ-

ent matters are being inquired about. Thus, as firmly grounded as the PDA appears to be in the observable behaviors of the job of firefighter, it does not appear that those observable behaviors provide much support for the analysis accomplished by the PAA.

The Test Development

As Dr. Fleishman testified and as AIR's reports to the City confirm, the principal reason for selecting the PAA method of job analysis for use in connection with Exam 3040 over the other techniques explored pursuant to the 1972 contracts was that the appropriate tests to determine applicants' abilities identified as necessary to the fireman's job had already been the subject of extensive study chiefly by Dr. Fleishman himself in his 1964 book and later work. As already noted, the appropriate test for each

ability, Dr. Fleishman had determined, was, all else being equal, the one most factorially pure for the abilities determined by the PAA to be important for the job, that is to say, tests with relatively high loadings on those factors and relative low loadings on other factors. In order to assure that the tests did not unduly overlap in measuring the same abilities, and also as the beginning of an effort, which could not draw to any substantial degree on Dr. Fleishman's prior work, namely, an effort to develop an appropriate scoring and weighting system, AIR, in December 1973, tried out an experimental battery of 11 factorially "pure" tests of eight physical abilities selected on the basis of the PAA, as well as a test of discrimination reaction

time, a psychomotor ability.[7] The 11 physical tests tried out in December 1973 and the eight abilities (in the rank order established by the PAA) they were intended to measure were as follows:

Ability	Test
1. Stamina	1. One Mile Run/Walk
	2. Five Minute Free-Style Stepping
2. Static Strength	3. Hand Grip Preferred
	4. Hand Grip Non-preferred
3. Explosive Strength	5. Free-Style Broad Jump

[7] Dr. Fleishman testified at trial that a group of 9 abilities identified by him, called psychomotor abilities, exists between the physical and mental abilities and can be tested for by tests which bear comparison at times with physical tests requiring special apparatus and at other times with the paper-and-pencil tests used to test mental abilities. The psychomotor test here discussed was a paper-and-pencil test.

4.	Gross Body Equilibrium	6.	Balancing
5.	Extent Flexibility	7.	Twist and Touch
6.	Gross Body Coordination	8.	Cable Jump
7.	Dynamic Strength	9.	Push-Ups
	Truck Strength, a subcategory of Dynamic Strength	10.	Leg Lifts
8.	Dynamic Flexibility	11.	Bend, Twist, and Touch

The tests were administered at the Baruch Recreational Center, which contains a small gym, on the Lower East Side of Manhattan, to 68 men then in training to become firefighters and to 32 firemen who had recently completed their training and were in the first probationary year.[8] All of the persons taking the

---

[8] Although a total of 100 men took the tryout test, only 99 results were used, because one trainee had a bad cough and was dropped.

trial tests had, thus, taken and passed the previous physical Exam 0159, administered by the Fire Department two years before, in 1971.[9] The reason for using trainees and probationers as test subjects was to try out the tests to be used on a sample closely approximating the applicant population expected to take Exam 3040. (The sample was not confined exclusively to trainees because there were not enough at the time to produce a test sample of 100.) Also, to this same end, and mindful of its obligations to the litigants in the Vulcan case, the City selected a sample designed to in-

---

[9] Examination 0159 had a five-part physical portion consisting of an agility test, a dumbbell lift (pectoral), a dumbbell lift (strength), a dumbbell lift (supine), and a broad jump. The exam was a qualifying exam, administered on a pass/fail basis, unlike Exam 3040.

clude a sufficient number of white, black, and Hispanic men to permit evaluation of the relative performance of each group on the various tests. Since, however, it was anticipated at this point that women would again not be permitted to apply for the firefighter's position, no women were included in the December 1973 tryout sample.

After the tryout battery was given, AIR computed correlations among the various tests to determine the degree of overlap between them and, also, correlated the tryout scores of 95 of the men in the sample with their scores on the five physical tests of Exam 0159 (the physical test previously used by the City), one of which was an agility test. The use to which the results of these correlations were put in determining which tests were to be included in Exam

3040 and, most importantly, in establishing the scoring of the exam will be discussed infra.

Before turning to the content of the battery of tests recommended by AIR to the City for inclusion in Exam 3040, however, it is appropriate to note at this point the occurrence of another event which was, in its implications, at least, to have profound effects on the test development of Exam 3040. In early 1974, because of concern about the impact of a height requirement on Hispanic persons, then-Fire Commissioner O'Hagan asked AIR to consider the appropriate height requirement for the fireman job.

The Height-Related Tests

On January 14, 1974, 12 individuals of varying heights (six firemen and six non-firemen from the Model Cities Program) performed a number of actual job

sample tasks, thought to be height- and job-related, to determine what relationship, if any, existed between the height of an individual and his ability to perform firefighting tasks.

Subsequently, on January 29, and February 2, 1974, the following tasks and one test, again thought to be height-related, were administered to a group of 100 firemen and 100 non-firemen, all male, including persons shorter than the City's then current 5'7" minimum height requirement for firefighters:

(1) Ladder Climb -- Subject climbs 35-foot ladder wearing a Scott air pack, coat, boots, and hat;

(2) Ladder Raise and Positioning -- Subject raises non-extended 20-foot length of 35-foot extension ladder from floor to wall and back down;

(3) Window Ventilation -- Subject swings six-foot hook to a designated target area in upper right quadrant of a wall;

(4) Hose Unclamping -- Subject removes one end of a hose section from a 74"-high clamp (with an examiner's assistant unclamping and steadying the other end), lowers the hose to the ground, and then raises and re-clamps;

(5) Victim Rescue -- Subject carries a 120-pound dummy up and down one flight of stairs one time;

(6) Vertical Hose Stretch -- Subject hauls over the shoulder the nozzle section of a 2-1/2" hose up and down one flight of stairs;

(7) Static Strength -- Hand grip test.[10]

---

[10] The basis on which these tasks and tests were determined to be height- and job-related so as to be included in the height study is not at all clear. Nor, with the exception of the hand grip test for static strength, does it appear that the tasks selected to be performed bore any relationship to AIR's job analysis. What is significant about the height study is that it appears to be the first instance of a problem which was eventually to compromise severely the rigor of AIR's test preparation, namely, a tendency on the part of AIR to defer in an entirely unsystematic way to the say-so of individual job incumbents when confronted with a claim that a particular test, derived from a source other than Fleishman's prior work, was job-related. As one example, the job

(Footnote continued)

On February 22, 1974, immediately following these tests, AIR sent to the City a recommended test plan for both the written and physical portions of Exam 3040, including recommended tests and scoring tables. For the physical test battery, AIR recommended eight tests to test the parenthetically indicated abilities: the one-mile run/walk (stamina), hand grip preferred (static strength), dummy carry (static strength, explosive strength, dynamic strength, and gross body equilibrium), agility (explosive strength, stamina, dynamic strength and

---

(Footnote 10 continued from previous page) task described in the height-related study as a victim rescue, accomplished by slinging a football blocking type dummy over one's shoulder and running up (and down) a flight of stairs, appears, on the basis of the testimony of fireman witnesses at trial, to be a highly uncharacteristic fireman's carry, if used at all.

gross body coordination), free-style broad jump (explosive strength), balance (gross body equilibrium), twist and touch (extent flexibility), and push-ups (dynamic strength).

Of these eight recommended tests, five were factorially pure tests discussed in Dr. Fleishman's 1964 book, and one, the mile run, was a factorially pure test for stamina not included in his book but recognized in the literature. However, as indicated by the descriptions given above, the agility test and the dummy carry were also recommended although they were not factorially pure tests. Their inclusion in the test battery requires separate discussion.

The "Face Valid" Tests

The agility test recommended in February 1974 -- consisting of an obstacle course with two walls, one 5'

and one 8' -- appears to have been drawn directly from the prior Fire Department Exam 0159, administered on a pass/fail basis in 1971. Why it was substituted, as it appears to have been, for the more factorially pure cable jump test is not at all clear.[11] Whereas the cable jump test is said by Dr. Fleishman to measure

---

[11] One likely explanation offered for the inclusion of the agility test is that the two groups identified as adversely impacted in the Vulcan litigation, namely, blacks and Hispanics, were perceived to do better on the agility test than whites. If so, inclusion of this test represented a step in the direction of decreasing the adverse impact of the physical test on blacks and Hispanics. Such a laudable purpose (while clearly having a bearing on the City's good faith) cannot excuse the lack of a rational connection between the test and the job analysis. Nor, indeed, would the addition of such a test, simply to reduce adverse impact on one class of employees, appear to comply with Judge Weinfeld's direction in the Vulcan litigation that the preparation of a replacement for the physical portion of Exam 0159 proceed in accordance with professionally accepted methods of test preparation.

the sixth-ranked ability of gross body coordination, the dodge-and-run aspect of the agility test recommended is said by Fleishman in his 1964 book to measure the third-ranked ability of explosive strength -- an ability already tested for by two of the recommended subtests (free-style broad jump and dummy carry). Moreover, in calculating intercorrelations between scores for the agility test on the December 1973 tryout and on Exam 0159, AIR determined that the agility test had a moderate correlation with tests for explosive strength and stamina, a somewhat lower correlation with gross body equilibrium, and the lowest correlation of all with gross body coordination -- the ability tested by the cable jump test for which it was substituted. This substitution became of crucial importance for plaintiff and the class she

represents because the height-related eight-foot wall which formed part of the agility test became not just a stumbling block, but a literal barrier for almost all women taking the exam.[12]

In January 1974, immediately before submission of the proposed test battery to the City, an effort was made by AIR to justify the inclusion of the eight-foot wall in the agility test. The AIR letter reporting the results of an effort to justify the height chosen, however, explains its rationale not in

---

[12] One woman who scaled the eight-foot wall -- not once, but twice -- testified she was able to do so by reason of a simple technique taught her in a private training school she attended, not apparently involving any of the abilities purportedly measured, namely, running up the wall with rubber-soled shoes before attempting to grasp the top rather than, as most women did, standing before the wall and jumping up.

terms of the height needed to establish the ability (explosive strength) sought to be tested, but rather in terms of the nature of the fireman's job. Moreover, rather than refer to the results of any of the job analyses already performed in order to determine if the analyses had established the relative importance of scaling walls as a task performed by firemen, AIR concluded, apparently on the basis of a single unrecorded interview between an AIR representative and representatives of the Fire Department, that firemen in New York have occasion to encounter eight-foot parapets between roof tops which must be scaled in the course of their work.[13] The relative

---

[13] The only evidence of such a characteristic of New York architecture introduced at trial was a Factory Exit Rule adopted by the New York City Board of Standards and Appeals in 1947 implementing the state's Labor Law by
(Footnote continued)

importance of this work behavior to the job as a whole is nowhere addressed in

---

(Footnote 13 continued from previous page)
requiring special safety exits to be provided for factory buildings, built before 1914 and then still in use, of five stories or less in height, where the distance between the roof of the building and the roof of the next building exceeds eight feet. Whether or not this provision can be taken as establishing the existence of the eight-foot differential as a common characteristic of factory buildings of this height and vintage at the time of enactment of the Code provision, the testimony at trial indicates that there are at present exceedingly few buildings in this category. The witness who testified for the City on this subject -- a fireman of over two decades' experience -- related that on the only occasion he has had to climb an eight-foot wall from one building to another he used the scuttle cover from a roof opening which he leaned against the side of the wall as a step. Other after-the-fact rationalizations were offered at trial for the eight-foot wall, such as the need to leap into the air to grab the lower rung of fire escape ladders said to hang on occasion eight feet or more above the street -- an ability not tested by the agility test actually used, since the technique taught and permitted to scale the wall in the test involved not leaping in place,
(Footnote continued)

AIR's job analysis.

The dummy test (as it came to be known) appears to have had its origin in the height study requested by then-Commissioner O'Hagan. Again the reason for inclusion of the test in the February

---

(Footnote 13 continued from previous page) but running up the wall as far as one could before grabbing its top. Still another fireman testified at trial to an occasional need to climb to the top of an enclosed exit to the roof of buildings to break through a skylight and create ventilation. However, as noted, the need to climb an eight-foot wall as a work behavior of sufficient importance to warrant its role in Exam 3040 does not appear to have been established on the basis of any of the systematic job analyses done. I conclude, accordingly, that the wall was included in Exam 3040 simply because of its inclusion in earlier tests and a subjective feeling on the part of City personnel that firemen should be able to do such things. As the EEOC Guidelines wisely inform us, such seat-of-the-pants justifications for employment tests may well be the source for the stereotyped images of jobs which exclude women without there being a rational reason for their exclusion.

recommendations is unclear. Whatever the explanation for the inclusion -- whether a desire to aid minorities or a recognition, which later proved accurate, that the City and the Department of Personnel would not accept too many tests with which they were unfamiliar either from earlier exams, the athletic field, or notions of what the fireman's job entailed -- inclusion of the test presented the same difficulties, given AIR's overall approach, as the inclusion of the agility test. Again, the complexity of the test assured that more than one ability was, in all likelihood, being tested for, thus increasing the problem, sought to be avoided by Fleishman's factorially pure approach, that scoring would be distorted by measuring more than once for the same ability. In addition, since neither the dummy carry nor the agility

test arose out of any systematic analysis of observable work behaviors (although appearing to owe their justification to such an origin), the addition of the tests reinforced the need for a criterion-related validity study.

The same movement away from the measurement of the relatively abstract abilities required for the job based on Fleishman's abilities analysis towards so-called "face valid"[14] tests continued

---

[14] As noted in the American Psychological Association, Inc. Standards for Education[al] and Psychological Tests ("APA Standards") at p. 29: "It should be clear that content validity is quite different from face validity. Content validity is determined by a set of operations[,] and one evaluates content validity by the thoroughness and care with which these operations have been conducted. In contrast, face validity is a judgment that the requirements of a test appear to be relevant. The writing of items in terms used in a particular job . . . may give an appearance of validity while contributing nothing to content validity or indeed to any other useful

(Footnote continued)

thereafter as City officials began to criticize and question the proposed test battery, in some cases on quite subjective grounds. Thus, both the Chief of the Fire Department and the Chief of the Department of Personnel complained of the scoring plan for the mile run (which permitted a candidate to complete the mile in 12 minutes and still not fail), not on the basis of its methodology, but simply because they or some close relative could run the mile in less time. Similarly, the so-called twist-and-touch test for the ability of extent flexibility and the balance test for gross body equilibrium were the subjects of complaints by then-Commissioner O'Hagan to

---

(Footnote 14 continued from previous page)
validity information (although such information may serve a useful public-relations function)."

representatives of the City Department of Personnel and through the Department to AIR on the ground that what was wanted was a test "for firefighters, not for ballet dancers."[15]

In response to this criticism of the mile run, AIR appears to have initially reacted quite appropriately by re-examining the data on the basis of which its scoring of the mile run had been determined. This re-examination revealed that AIR had developed the March 1974 scoring norms based on the distribution of scores in the December 1973 tryouts, using a 31-lap track at the Baruch gym. Obviously, the effect of having so

---

[15] Ironically, among the relatively neutral job anchors used by Dr. Fleishman in his job analysis, the abilities of ballet dancers were recognized as one reliable indice of human abilities common to ballet dancers and firefighters.

many short laps was to increase the average time of the runners. When City officials complained that the scoring system was too lenient and insisted it be made more stringent, AIR asked to try out the mile run again on a four-lap track. This request was, however, rejected. Instead, the representatives of the City's Department of Personnel directed AIR's attention to a standard reference work by T.K. Cureton, Jr., which set out suggested scores for the mile run based on norms developed from runs performed by an all-male sample population consisting of high school, college, and military men involved in a program of daily physical training in preparation for military

service in World War II.[16] AIR accepted the suggestion that this more stringent scoring norm be used with little if any attention to the logic of using such norms in its overall scoring system, either in terms of job requirements or norms in the applicant pool.

A substitute for the rejected twist-and-touch test appears to have emerged from a visit by AIR, Fire Department personnel, and a representative of the Personnel Department to a physical

---

[16] Even so, the scoring for the mile run, as it was finally established for the physical portion of Exam 3040, was such that only 50% of the trained Cureton sample would have been able to achieve a passing score despite the fact that the Exam 3040 mile was run on a ten-lap, rather than a four-lap, track. Based on their recorded times in the tryouts on a 31-lap track, none of the incumbent firemen who tried out the one-mile run in December 1973 would have received a passing score on Exam 3040.

examination for firefighters being administered by the City of Chicago in August 1974, made at the specific request of then-Commissioner O'Hagan who had already expressed his dissatisfaction with AIR's test battery. The Chicago battery consisted of five tests: flexed arm hang, obstacle run, stair climb, man lift and carry, and hose coupling. The AIR representative summoned to Chicago was asked to observe and evaluate the Chicago battery as compared with AIR's.

Following this visit, AIR continued to recommend the tests it had originally proposed, based primarily on differences in the job analysis and testing techniques being used, with Chicago testing for multiple abilities in tests resembling actual job tasks and New York testing for single abilities by means of factorially pure tests highly loaded for

a single factor deemed important to the job. In addition, AIR opposed use of the one factorially pure test used in Chicago, the flexed arm hang, because of the low degree of reliability of that test from one administration of the test to another -- so called test/re-test reliability.

Not satisfied with this response, a representative of the City's Department of Personnel then contacted, apparently without consulting AIR, the consultants used by the City of Chicago in preparing that City's physical exam and asked those consultants to criticize AIR's proposed test battery. This criticism served to confirm the impression of the City officials that the mile run scoring was too lenient and that the twist-and-touch and balance tests were

"inappropriate" for firemen.[17] In addition, the Chicago experts pointed out that at least one of the tests, push-ups, would have an adverse impact on women because of their lower average upper-body strength.[18]

Without reaching any professionally acceptable resolution of these differences, AIR thereafter acceded to

---

[17] The Chicago experts' complaint about the twist-and-touch test was somewhat more reasoned than Commissioner O'Hagan's. Thus, they are reported to have noted that firemen must "twist and touch," not just to one side of the body, but to both sides. This criticism, while rational in terms of a test based on actual job behavior, misconceived the purpose of the AIR test, which was to determine the abstract ability of extent flexibility which Dr. Fleishman's studies showed could be evaluated by measuring the candidate's ability to twist and touch in one direction only.

[18] In addition, the push-up test was reasonably criticized as difficult to administer and monitor.

the City's requests by substituting a test dubbed a "ledge walk" test for Fleishman's balance test of gross body equilibrium, adding a new "obstacle" to the Exam 0159 obstacle test called "window-ladder-window" as a test for extent flexibility and replacing push-ups with what AIR had previously termed a relatively unreliable test of dynamic strength: the flexed arm hang. Both the "window-ladder-window" obstacle and the flexed arm hang were drawn directly from the Chicago battery. The "ledge walk" appears to be entirely unstudied and unprecedented as a test of the ability of gross body equilibrium which it purported to measure either in terms of factor analysis or reliability. The "window-ladder-window" test, by being incorporated into the agility test, came to measure extent flexibility (to the extent

it can be said to have reliably measured it at all) in terms of time rather than in terms of the inches to which one could extend oneself, as under the original twist and touch test. The ledge walk, performed with eyes open, no longer eliminated that type of balance produced by visual eye contacts, as was the case with Fleishman's balance test. In fact, there appears no basis for saying that any of the three tests reliably measured the abstract abilities measured by the Fleishman tests for which they were substituted.

As the result of these additions to the proposed test battery, it became necessary to conduct another tryout test to establish new scoring norms. This tryout was administered in early November 1974 to a group of 75 incumbent firemen, only six of whom were probation-

ers. The November tryout battery consisted of the new tests: the flexed arm hang, ledge walk, and a window-ladder-window test (consisting of crawling through a window, across a horizontal ladder and through another window). In addition, the dummy carry was tried out again apparently in order to get new scoring norms.

A final proposed test battery for Exam 3040 was forwarded to the Department of Personnel on November 29, 1974. The physical portion of the exam contained seven physical subtests. The test and the abilities they were said by AIR to measure were the mile run (stamina), hand grip (static strength), broad jump (explosive stength), dummy carry (static strength, explosive strength, dynamic strength, and gross body equilibrium), agility (including window-ladder-

window) (explosive strength, stamina, dynamic strength, gross body coordination and extent flexibility), ledge balance (gross body equilibrium), and flexed arm hang (dynamic strength).

Scoring

Because AIR's development of test scores did not arise out of Fleishman's prior work, but rather had its origin at least in part (leaving to one side the scoring of the mile run already discussed) in empirical studies conducted during the test development, the subject merits separate discussion.

As already noted, the December 1973 tryouts were designed to provide AIR with the basis for a norm-referenced scoring system based on a sample that was thought to approximate the applicant population. To this end trainees were used to the extent they were available

with probationary firemen filling out the complement of persons tested. No women were included in the sample intended to represent the applicant population; and the sample, by definition, differed from the applicant population in at least one other notable characteristic, namely, that all members of it had passed (on the pass/fail basis on which it was administered) Exam 0159.

In order to get norms for the tests added subsequent to the December 1973 tryouts, however, AIR appears to have departed to some extent from its intention of using an applicant sample. Thus, norms for the dummy carry were drawn from the height-related tasks tryout on firemen and non-firemen held on January 29 and February 2, 1974. (Although the hand grip test was also administered to the same group, AIR used its

December 1973 based sample perhaps on the ground it more closely resembled the applicant population.) On the other hand, to obtain norms for the agility test drawn from Exam 0159, AIR turned to the test scores reported for the exam when it was administered in 1971, thereby securing a sample resembling perhaps as closely as possible (but for the exclusion of women) a sample representative of what might be expected to turn out for Exam 3040.

As noted, the agility test had been modified to include a window-ladder-window component. Scores on the window-ladder-window component from the November 7 and 8 tryout were added by AIR to the original distribution of scores on the agility test derived from Exam 0159 by re-working the original 1971 distribution on the assumption that those persons

who did well on window-ladder-window would also have done well on the agility test in its previous form -- an illogical assumption if the window-ladder-window component was supposed to test for the discrete ability of extent flexibility, an ability not included in the former agility test.

In setting the scores for the flexed arm hang, although AIR had data both for the 75 firemen who tried out for this test on November 7 and 8, 1974, and data from the arm hang test in Chicago, AIR disregarded the New York City scores and set the scoring table according to the Chicago data. This made some sense in terms of a desire to replicate an applicant population, since the November 1974 group consisted of 75 incumbent firemen, only six of whom were probationers. However, one paradoxical result of

this choice was that a large number of the New York City firemen who performed the flexed arm hang as administered in November 1974 would have scored at or below zero on this portion of Physical Exam 3040, as scored based on the Chicago norms.

In the scoring of the ledge balance, for which there was no normative data other than the November 7 and 8, 1974 tryouts, that data was necessarily used. However, as a result of using as norms the performance on the test by a group of job incumbents rather than performance by an applicant pool, the scores were set at a level at which many of the New York City firemen who tried out the test would have received an unacceptably low score.

This result came about because AIR applied to all scores -- whether for

groups of job incumbents or for groups purporting to reflect the applicant pool -- traditional grading techniques, fixing the pass/fail levels at what appeared to be "natural breaks" in a roughly drawn curve in the distribution of scores. AIR then, also using traditional grading techniques, fixed 70.0% as the appropriate designation for a passing score, thereby effectively dividing those passing the exam into 300 separate ranks, each representing a tenth of a percentage point between 70.0 and 100.0%. Selection of candidates from within each of these 300 ranks (in cases in which a number of candidates scored at the same percentage level) was made on the basis of candidates' social security numbers -- that is, essentially on a random basis.

On March 6, 1974, AIR sent to the City a proposed scoring table for the

recommended battery of eight tests. In determining the proposed tables, AIR first determined the relative weighting of each test based on the rankings of the mean ratings of the physical abilities in the PAA. For example, because stamina received the highest mean rating on the PAA, AIR proposed that the mile run, which tested stamina, should be weighted 20 percent of the final score or 200 points out of a total of 1,000 points. (1,000 points was to equal a score of 100%.) The other five recommended factorially pure tests were to be given a weight of 100 points each. In order to deal with what had become known as the mixed ability tests, it was proposed that the agility test and the dummy carry were to receive 150 points each. Thereafter, when one of the factorially pure tests -- the balance test -- was incorporated into

the agility test by the addition of the window-ladder-window component, the 100 points attributable to the balance test was simply added in toto to the agility test with the paradoxical effect that one of the mixed ability test[s] received 150, and the other 250, points, becoming the most important test of all. Another paradoxical result of this system of weighting was that the mile run, said to test for the ability rated first as a needed ability for firefighters, was assigned a total of 200 points while the hand grip test, a measure of static strength, the next highest rated ability, was assigned a total of 100 points.

At the same time as representatives of the City expressed themselves not satisfied with the proposed test battery, they also complained about the recommended scoring tables for those

tests. Not only was the scoring for the mile run criticized, as noted above, but questions were also raised on a not much more objective basis concerning the scoring for the broad jump, the dummy carry, and the hand grip, AIR responded to these criticisms by "tightening" the proposed scores on each of them. Thus, with regard to the broad jump, the original proposal gave no credit on the test unless the candidate jumped at least six feet. At the insistence of the City, this minimum distance was raised to 6 foot 2 inches. With regard to the hand grip, the minimum pressure required to obtain the minimum score on the test was raised from between 28 to 30 kilograms to between 34 to 35 kilograms. There appears no consistency in these increases in the minimum scores (which served to deprive some women candidates of any

score at all on the subtests -- 5 out of the total 79 on the broad jump and 3 on the hand grip) nor any plausible reason for them, apart from acquiescence in the City's view that the scores provided were too lenient.

Finally, it must be noted that for reasons not explained by the evidence at trial AIR used no score conversions in arriving at a combined score for the seven tests comprising the physical examination and for the written examination. The effect of this was to give a far greater weight to the physical portion of Exam 3040 than was intended. The reason for this is as follows. As already noted, Exam 3040 consisted of both a written and a physical part, each weighted 50 percent in determining the candidate's rank order. In any test consisting of more than one part (each to

be given equal weight), it is necessary to insure that one test not become, de facto, the determinative one because of greater variance in results on one test than the other. The generally accepted method of avoiding this result is to convert the raw scores on each test into a standard unit before the results of the test are combined to determine the candidate's overall rating. Use of such a standard score would have insured that the variance in the two sets of raw scores was taken into account so as to maintain an effective 50/50 rating.

Test Administration and Results

As noted above, Exam 3040 included a written test and a physical test. The written test consisted of 100 short-answer questions. Only those candidates who passed the written test were allowed to take the physical test bat-

tery. As noted above, candidates who passed the physical test received a final combined score consisting of the average of their scores on the written and physical tests. Candidates who were veterans received an additional five points, which were added to their combined score.

The written test was given on December 3, 1977, to 25,168 persons: 24,758 males and 410 females. (It is these 410 women who constitute the class, as presently defined.) A total of 24,252 males and 389 females passed the written test. The physical exam was administered over the period February 15, 1978, to April 30, 1978, to a total of 18,148 persons: 18,060 males and 88 females. Of these, 16,925 males and 79 females completed the physical test. A total of 7,847 males and no females passed the test.

Prior to administering the test, as part of an effort stimulated by the Vulcan litigation to recruit minorities, an effort was made by the City through the City University system to encourage women to apply for the job of firefighter and take the test. Another effective means of recruitment (to judge from the witnesses who testified at trial) was the encouragement by incumbent firemen of female family members (sisters, nieces, and daughters) to take the exam.

All applicants completing a formal application form were sent a 104-page booklet which described the written and physical tests. The booklet included a description of each physical subtest as well as a description, with pictures depicting both men and women, of 25 preparatory exercises and general informa-

tion on physical fitness. In addition, the City conducted a "physical fitness familiarization program" in anticipation of the physical portion of Exam 3040. Four sessions were held outdoors at the Fire Department's division of training at Wards Island during December 1977, and the remaining eight sessions were held indoors during January 1978 at the Sum-[n]er Avenue Armory in Brooklyn, the site of the actual test.

More important than this "familiarization" program in terms of test preparation appears to have been an informal system of schooling conducted privately for a fee by City firemen on their own time. While the familiarization program succeeded in giving candidates a concrete idea of what the physical tests would look like, only the private sessions gave candidates an opportu-

nity to practice and train for the physical tests. As established by the testimony at trial, including that of one woman candidate who scored second among the women on the test, these training and practice sessions were of enormous value in teaching techniques of mastering the test, particularly with the so-called multiple ability tests -- the dummy test and the agility test (with its eight-foot wall) -- and the ledge walk.

The physical exam as administered consisted of the following tests: dummy carry, hand grip, broad jump, flexed arm hang, agility test, ledge walk, and one-mile run, administered in this order. No particular order to the administration of the test appears to have been prescribed by AIR. What follows describes the administration of each of the tests.

1. Dummy Carry. The test was administered and scored as described to the candidates in the booklet provided them as follows:

Dummy Carry Test Weight = 150

Standing with both feet behind the start line, on the signal "Go," step forward, pick up the 120-pound dummy, put it on one shoulder, carry it in that position up one flight of stairs, walk around the marked lines of the upper stair landing, return down the stairs, bend down on one knee, then place the dummy back on the mat under control. The dummy must be kept off the ground when carrying it. Score is time to completion.

A two-second penalty is added to the total time score if the dummy is "dropped" out of control after the dummy has been carried up and down the flight of stairs. "Dropping" the dummy anywhere else in the test, such as on the stairs or the upper stairs landing results in a zero score.

One trial is to be given unless a candidate receives a zero penalty score, in which case a second and final trial is allowed.

Score (in seconds)	Weighted Percent
18 or less	150
19	145
20	141
21	137
22	133
23	129
24	125
25	121
26	117
27	113
28	109
29	105
30	100
31	95
32	90
33	85
34	80
35	75

Still unfinished after 35 seconds:

Lowering dummy to ground: 70

On upper stair landing or returning down stairs: 35

Can't lift or going up stairs: 0

As already noted, the dummy test was said by AIR to test for static strength, explosive strength, dynamic strength, and gross body equilibrium.

The dummy used in the test was cylindrical, covered in canvas, without

handles or other articulation, and presented initially a considerable technical problem to the uninitiated as to how to lift and hold it. Anyone unable (whether through failure of technique or a failure in the capacities of strength needed to lift the dummy) lost not only his or her score for the abilities said to be involved in lifting, but also any score on the others of the multiple abilities sought to be tested during the test, including gross body equilibrium, ranked number four among the abilities needed for the job. Out of the 80 women taking the exam, only four appear to have succeeded in lifting the dummy.

2. The Hand Grip. The requirements and scoring for the hand grip test were as follows:

Hand Grip Test Weight = 100

Holding the dynamometer in the preferred hand, with the palm around the top bar and the fingers curled around the bottom bar, arm extended downward away from the side of the body and with the palm facing the side of the body, squeeze the dynamometer as sharply and as steadily as possible. This test is to be performed while standing. Score is the highest dynamometer reading obtained on one squeeze.

One squeeze is a trial. Two trials, separated by at least one minute, are to be given. The best of the two trials is to be rated.

Score (in kilograms; kg. = 2.2 lbs.)	Weighted Percent
70 or more	100
65 - 69	95
60 - 64	90
55 - 59	85
50 - 54	80
46 - 49	75
42 - 45	70
40 - 41	65
38 - 39	60
36 - 37	55
34 - 35	50
Less than 34 kgs.	0

The ability said to be measured by the hand grip is static strength. The principal difficulty in administration of

the hand grip appears to have been that, although the dynamometer employed contained a means of adjustment for hand size, AIR did not explain its use to the City, and the City's representatives did not explain its use to candidates. As a result, plaintiff, for one, found difficulty in getting a grip on the instrument with, undoubtedly, some significant effect on her score. It may be inferred, moreover, that there was as a result of this error in administration an adverse effect on women in general as a result of smaller hand size on average.

3. The Free-Style Broad Jump. The broad jump requirements and scoring were as follows:

Free-Style
Broad Jump — Test Weight = 100

With both feet behind the start line, jump forward as far as possible without falling backwards. Arms may be used free-style during a jump. Score is greatest distance jumped as mea-

sured from the point where the rearmost heel lands to the start line.

Falling backward when landing results in a zero score for that trial. The best of two trials is to be rated.

Score (in feet and inches)	Weighted Percent
8'08" or better	100
8'05" or better	95
8'02" or better	90
7'10" or better	85
7'06" or better	80
7'02" or better	75
6'10" or better	70
6'08" or better	65
6'06" or better	60
6'04" or better	55
6'02" or better	50
Less than 6'02"	0

The test was said to measure (as did the dummy carry) explosive strength.

One problem encountered by plaintiff in the administration of the test was that at the time she took the test it was administered on a mat which had a tendency to slip on the floor. The mat slipped in her case, and she fell backwards as a result. The mat was later

removed. These and other minor difficulties in test administration are complained of by plaintiff not because they affected women only but because they call into question the degree of precision in scoring with which the test purported to measure differences in the abilities of candidates to perform on the job.

4. Flexed Arm Hang. The test's requirements and scoring were as follows:

Flexed Arm Hang Test Weight = 100

Mount the ladder to a height such that the chin is even with the horizontal bar; use a firm overhead grip on the bar and, when ready, nod; the ladder will be taken away and the timing will begin. Hang onto the bar as long as possible. The test will be over when arms are completely extended. Score is time before arms are completely extended.

One trial is to be given.

Score (in minutes and seconds)	Weighted Percent
2'20" or more	100
2'10" - 2'19"	95
2'00" - 2'09"	90
1'50" - 1'59"	85
1'40" - 1'49"	80
1'30" - 1'39"	75
1'20" - 1'29"	70
1'15" - 1'19"	65
1'10" - 1'14"	60
1'05" - 1'09"	55
1'00" - 1'04"	50
Less than 1'00"	0

The test was intended to measure for dynamic strength (also measured by the dummy carry).

One notable difficulty in the administration of the flexed arm hang was, as AIR noted in recommending against its use by the City, its low (.53) test/re-test reliability. This low degree of reliability was attributed to difficulties in obtaining objective measurements because of subjective differences between monitors in deciding when the arms are "completely extended."

5. Agility Test. The agility test was administered and scored as follows:

Agility Test Weight = 250

From a position lying on the back, feet together, hands at side, on signal "Go," rise and run nine feet to the five-foot wall and scale it; run 11-1/2 feet to a maze of obstacles and dodge through; run about 15 feet and crawl through the first "Window;" mount the 30-foot horizontal ladder and crawl across it, keeping one hand on each outside rail; crawl through the second window; run about 20 feet to the eight-foot wall and scale it; run around one obstacle; sprint 45 feet back to the finish line. Score is time to complete the course.

A penalty of 2 seconds will be added to the total time score each time one foot touches the floor on the outside of the ladder while crawling across it. A penalty of two seconds will be added to the total time score if both feet fail to land beyond the last rung of the ladder and on the side of the ladder farthest from the second window.

Use of the iron supporting rods on the eight-foot wall to aid the climb or running out of the course without retracking and continuing properly within the time limit will result in credit only for the previous properly completed obstacles. The best of two trials is to be rated.

Score (in seconds)	Weighted Percent
22 or less	250
23	245
24	240
25	235
26	229
27	223
28	217
29	211
30	205
31	199
32	193
33	187
34	181
35	175
36	169
37	163
38	157
39	151
40	145
41	139
42	132
43	125
Still unfinished after 43 seconds:	
Sprint	118
Eight-foot wall	59
Window, ladder, maze, or five-foot wall	0

The test was said to test for explosive strength, stamina, dynamic strength, gross body coordination, and

extent flexibility. Almost all of the women candidates appear to have been unable to clear the eight-foot wall. Whether this was because of a failure of technique or because of lack of ability cannot be determined in individual cases. However, technique clearly played a large part in the successful completion of the test by other candidates. Since the eight-foot wall appeared at the end of the obstacle test and since failure to clear it automatically dropped one's score from a possible 250 to 59, a candidate who might have scored well on all of the other abilities tested for (including gross body coordination, which was tested only by the agility test and which was deemed by AIR to be worth 100 points when tested separately by its balance test) was effectively deprived, by means of a failure on the eight-foot wall, of accu-

rately demonstrating abilities deemed by AIR to be necessary for the job.

6. Ledge Walk. The test and its scoring were as follows:

Ledge Walk Test Weight = 100

Wearing an "oxygen unit" consisting of a harness and weight on the back, for a total weight of 26 lbs., start with both feet facing the wall, within the one-foot area marked off on the 30-foot "ledge" that is two and one-half inches wide, and six inches from the wall. On the signal "Go," side step to the right as quickly as possible to the opposite side of the ledge without falling off; touch the side board with the right foot and make sure that both feet are within the one-foot marked-off area; then side step to the left as quickly as possible, but without falling off, until both feet are again within the one-foot marked-off starting area. Score is time to move across the ledge and back again without falling off.

"Falling" off the ledge constitutes a zero score for that trial. Two trials are to be given and the best trial rated.

Score (in seconds)	Weighted Percentage
9 or less	100
10	97
11	94
12	90
13	86
14	82
15	78
16	74
17	70
18	65
19	60
20	55
21	50
More than 21	0

The ledge balance was said to test for gross body equilibrium. One problem encountered by plaintiff in the administration of the test was that she was unable to adjust the harness for the breathing unit to her back and was not assisted by anyone in doing so. In addition, when plaintiff took the test one of the walls used leaned towards the candidates, making it more difficult to accomplish the walk. Again, this difficulty is mentioned not because it adversely

affected the ability of women to perform on the test -- it affected all equally -- but to illustrate the difficulties, discussed infra, created by attributing too great precision to the test instrument in predicting job performance.

7. Mile run. The test was administered in general as described for the candidates in the booklet provided them as follows:

One Mile Test Weight = 200

With both feet behind the start line, on the signal "Go," complete X number of laps around the track as fast as possible by running or a combination of running and walking. Score is time to complete X number of laps which is equal to one mile.

One trial is to be given.

Score (in minutes and seconds)	Weighted Percent
Less than 4'50"	200
4'50" - 4'56"	195
4'57" - 5'03"	190
5'04" - 5'11"	185
5'12" - 5'19"	180
5'20" - 5'27"	175
5'28" - 5'36"	170

5'37" - 5'45"	165
5'46" - 5'54"	160
5'55" - 6'03"	155
6'04" - 6'12"	150
6'13" - 6'21"	145
6'22" - 6'30"	140
6'31" - 6'39"	135
6'40" - 6'48"	130
6'[4]9"-6'56"	125
6'57" - 7'04"	120
7'05" - 7'11"	115
7'12" - 7'18"	110
7'19" - 7'24"	105
7'25" - 7'30"	100
More than 7'30"	0

The test was designed to measure stamina or aerobic capacity -- something it will do only for runners taught to pace themselves in running. In the absence of pacing, the test measures not stamina or aerobic capacity -- that is, the ability of the body to generate energy from oxygen uptake -- but anaerobic capacity -- the ability of the body to generate energy from itself, something also measured by strength tests. For this reason, a more generally accepted test for stamina than the mile run is a

test requiring the candidate to run or walk for twelve minutes, measuring the distance accomplished, on the theory that this test forces pacing on even an untrained candidate. Several of the candidates testified that they had trained themselves in pacing and that only this training had enabled them to complete the mile. The absence of an accurate test of stamina or aerobic capacity is said by plaintiff to have been of great significance in terms of the validity of the test because of the general consensus among the witnesses, both experts and job incumbents, that stamina rather than brute strength is the prime requirement for the job of firefighting. "Firefighting is," as one witness testified, "not a sprint event."

Adverse Impact

25,168 candidates filed applications and appeared for the written portion of Exam 3040, of which 24,758 were men and 410 were women. 506 men failed the written exam, and 21 women did so. A total of 18,148 persons who passed the written exam appeared to take the physical exam: 18,060 males and 88 females. Thus, over one-quarter (26%) of the men who passed the written exam (24,252) did not present themselves for the physical (6,192). Close to 77% of the women who took and passed the written exam (389) did not appear to take the physical exam (301).

Of the 17,004 candidates who completed the physical exam, 16,925 were men and 79 were women. 7,847, or 46%, of the men passed. None of the women did.

Data from the United States Bureau of Census, based on the 1970 census, relating to the Standard New York Metropolitan Statistical Area and to Orange and Putman Counties, show that 872,284 males and 1,014,706 females of the age group eligible for firefighters resided in the geographical area within which eligible persons are required to reside at the time of appointment. Expressed another way, 46.2% of this population was male, and 52.8% of this population was female. The total labor force between the ages of 18 and 29 residing in the identified geographic area was 655,022 males and 518,301 females, or 55.8% males and 44.2% females.

Roughly comparable data from the United States Bureau of Labor Statistics for the year 1979 indicate that in that year males constituted 55.3% and

females, 44.7% of the civilian non-institutional labor force in the age group and localities from which the applicants for Exam 3040 were drawn.

An eligibility list consisting of some 8,018 names in rank order was produced from Exam 3040. As noted above, each "rank" on the list consists of a tenth of a percentage point; and, where there are several candidates in the same rank, the call-up of candidates is made by random selection, using candidates' social security numbers. As of February 19, 1982, 2,666 men have been appointed from the list. Fifty places have been reserved for members of the plaintiff class, pursuant to an agreement between counsel, with back pay, benefits, and seniority as of November 5, 1980, should this Court determine that Exam 3040 was discriminatory and that members

of the plaintiff class are entitled to appointment to the Fire Department. The eligibility list expires June 21, 1982.

Discussion

I

The Validity of Exam 3040

As noted in Guardians Association v. Civil Service Commission, 630 F.2d 79, 88 (2d Cir. 1980) (Guardians IV):

> ". . . the accepted procedure for Title VII cases is to require the plaintiffs to establish a prima facie case, and then to require the defendants to rebut this showing with proof that the test was legitimately job-related. See Albemarle Paper Co. v. Moody, 422 U.S. 405, 95 S. Ct. 2362, 45 L. Ed. 2d 280 (1975); McDonnell Douglas Corp. v. Green, 411 U.S. 792, 93 S. Ct. 1817, 36 L. Ed. 2d 668 (1973); Griggs v. Duke Power Co., 401 U.S. 424, 91 S. Ct. 849, 28 L. Ed. 2d 158 (1971)."

A. The Prima Facie Case

In this case, as in Guardians IV, the exam at issue had a disparate impact by any reasonable measure including the standards developed by the Supreme Court in Castaneda v. Partida, 430 U.S. 482 (1977), and by the Equal Employment Opportunity Commission ("EEOC") in its Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607 (1978) ("Guidelines," hereinafter cited only by the subdivision numbers of 29 C.F.R. § 1607). Under Castaneda, in cases such as this involving substantial samples, "if the difference between the expected value [from the random selection] and the observed number is greater than two or three standard deviations," a prima facie case of discriminatory impact is established. 430 U.S. at 497 n.17. Under the Guidelines, "[a] selection rate for any

race, sex, or ethnic group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact." Section 4(D).

Here, the 0% pass rate for women is obviously less than 80% of the 46% pass rate for men, employing the test referred to in the Guidelines. The pass rates of the men and the women were separated from the results to be expected from a sex-neutral selection process by more than eight standard deviations, representing far more than the .03% likelihood of chance factors having produced the outcome found to establish a prima facie case of disparate impact in Castanada. Comparing the zero pass rate with the available work force figures

produces an even stronger case for non-chance selection factors.[19] In this case, the inference of discriminatory impact from facially neutral employment practices arises from "the inexorable zero." International Brotherhood of Teamsters v. United States, 431 U.S. 324, 342 n.23 (1977).

---

[19] Defendants question the failure of plaintiff to present evidence with regard to the number of men and women in the work force who have valid New York driver's licenses and meet the vision requirement of the N.Y.C. Admin. Code of 20/40 for each eye separately, without glasses, and the height requirement. For this reason, as well as the other unusual demands of the firefighter's vocation, it seems appropriate in determining adverse impact and relief, as noted infra, to rely on a comparison of the ratio of male to female pass rates to the male/female ratio among those actually presenting themselves for the physical examination rather than a comparison of the pass-rate ratio to the ratio of men to women in the work force. But cf. Dothard v. Rawlinson, 433 U.S. 321, 330 (1977).

Defendants argue that statistics alone should not be sufficient to establish a prima facie case of disparate impact here in the absence of proof to rebut another available inference from the statistics, namely, that none of the women who took the test were strong enough to do the work required by the job. This argument seeks to re-define the issues and reverse the burden of proof established by the Supreme Court for Title VII cases, see Griggs v. Duke Power Co., supra; Dothard v. Rawlinson, supra, by requiring plaintiff to show that she is an exception to the general characteristics of her sex and to prove that individual members of her sex are worthy of employment. Title VII imposes neither requirement on the plaintiff.

B. Test Validation

The issue in this case, as in Guardians IV, therefore, is whether defendants have rebutted plaintiff's prima facie case by showing that Exam 3040 was job-related, that is, that the test was valid because it accurately selected those applicants who will make better firefighters. Guardians IV, supra, 630 F.2d at 88.

The "threshold task" in determining the validity or job-relatedness of the test is to select the appropriate method of assessing its validity. Id. at 91. In this case defendants consistently, until almost the conclusion of the trial, asserted that only one technique of the three specified in Guidelines §§ 5(B) and 14 was being urged, namely, content validation. Accordingly, that proposition will be considered

first. The belated assertion that the test is criterion valid will be considered infra.

The proposition that the physical portion of Exam 3040 is content valid must be rejected. Defendants themselves and their outside consultants did not consider it to be such until the validity of their test was attacked. In all events "the abilities that the test attempts to measure are . . . [not] the most observable abilities of significance to the particular job in question." Guardians IV, supra, 630 F.2d at 93.

That Dr. Fleishman's nine categories of physical ability constitute anything other than constructs would seem beyond debate by almost any definition, whether it be the somewhat mechanical fact/inference distinction of Guidelines § 14(C)(2) or the more realistic view of

Guardians IV, which distinguishes between job content and job construct as "simply different segments along a continuum reflecting a person's capacity to perform various categories of tasks." Id. at 93. In terms of the purpose of the distinction between content and construct -- namely, the method by which one is going to determine on a firm, factual foundation that one has tested for the abilities or qualities needed for the job -- it is clear that Dr. Fleishman's technique can produce such factually grounded confidence only by a criterion study characteristic of construct validation. Nothing in the concepts of dynamic strength, gross body equilibrium, stamina, and the like has such a grounding in observable behavior or the way firefighters operate that one can say with confidence that a person who possesses a

high degree of these abilities as opposed to others will perform well on the job.

AIR's own description of the method of validation of the physical tests it proposed to develop through the PAA confirms that criterion rather than content validation is appropriate for a physical test grounded on the PAA. This description was in marked contrast to its description of the method of validation of the written exam, which was stated explicitly to be content validation. The proposal for the use of the PAA makes clear that criterion or construct validation was contemplated by AIR from the beginning of its project to be abandoned only when the City declined to fund the contemplated criterion study because of an asserted shortage of funds. The only mention of content validation appears in a 1974 progress report on the project

which, after noting the clear suitability of the PAA method of analysis for construct or criterion validation, states that the suitability of the PAA for content validation "depends" on the grounding of the PAA in a thorough task analysis. Here, the abilities analysis arose not out of an analysis of firefighting activities, but out of Dr. Fleishman's 1964 factor analysis of physical tests. While some effort was made to connect the abilities analysis to specific examples of firefighters' behavior, the effort can hardly be called thorough. Many of the concrete instances of the abstract abilities identified as important to the job are described with almost as great generality as the ability itself, e.g., "functions at fire operation." Other instances betray confusion concerning the content of the ability and create severe

doubts as to whether the job analysis method was understood and correctly applied, e.g., as an instance of dynamic strength, "have a fire late in a work shift." Even where concrete instances of work behavior as evidence of a particular ability are given, there is no systematic study of what the rater has in mind or of its relative importance or criticality. See Guidelines § 14(D)(2), and cf. Guardians IV, supra, 630 F.2d at 95.

What has been said might, on first impression, seem appropriately limited only to those factorially pure tests which remained part of the test battery and to have less application with respect to the so-called "face valid" tests: the ledge walk, dummy carry, and agility tests. Yet, in fact, the same conclusion applies to these tests as well. All of the tests were conceived of

and justified by AIR as tests for the factorially pure abilities identified by Fleishman in his 1964 work, not in terms of the job behaviors they mimic or appear to resemble. The multi-ability tests were explicitly referred to by AIR as "face valid," a term of art developed by the American Psychological Association, Inc., indicating that no pretense was made that these tests were content valid, that is, drawn from an assessment "of the important work behavior(s) required for successful performance and their relative importance." Guidelines § 14(C)(2). As APA Standards at p. 29 caution, while the concept of face validity may have some public relations role in selling a test to job applicants, face validity is no substitute for validation based on concrete analysis either of the job's requirements or the test's results.

Thus, while it is possible to find, scattered through AIR's job analysis, isolated references to work behaviors bearing superficial resemblance to certain of the face valid tests (although there is little or nothing about walking on building ledges, carrying people on one's shoulder or dodging and running through obstacles at top speed), the work behaviors reported are simply given as examples of the more generalized abilities which are at the center of AIR's study; they are hardly described with the care required to establish with certainty what is being referred to. Nor is there any indication of their relative importance to the job. As the evidence at trial established, a carry of the type illustrated by the dummy carry portion of Exam 3040 would be dangerous in the extreme to carrier and carried alike; ledge

walking of the type illustrated in Exam 3040 would be an extraordinarily rare event as an actual job behavior; and the top speed obstacle run bears little relation to the paced conservative approach, which, all witnesses agreed, is appropriate to effective performance at a fire. Finally, even if these techniques did represent actual job behaviors, it is clear from the demonstrated advantage possessed by those who were taught the special tricks by which the dummy is to be lifted and the eight-foot wall scaled, that a large part of the abilities measured represented those which applicants could be trained to acquire. As the Guidelines sensibly conclude, tricks of the trade that can be taught do not represent appropriate abilities to measure to determine whether the applicant can

perform successfully on the job. Guidelines § 14(C)(1).

To say that Dr. Fleishman's abilities analysis lacks content validity is, of course, not to criticize it as an unworkable job analysis device, but only to point out what Fleishman himself recognized, namely, the need for construct or criterion validation to see if the test in fact succeeded in predicting those persons who would perform the job best. Unfortunately, not only was the original criterion validation proposed a victim of the City's fiscal crisis, but a more limited concurrent validation study in which the predictive value of the test would be studied by its administration to job incumbents was also rejected. In the absence of either a content or criterion validation of AIR's analysis, the value of it remains a moot question. Moreover,

because of errors in the test preparation and scoring of the exam, it seems likely that even a belated criterion validation of Exam 3040 would be of little assistance in determining if those who did well on it would make better firefighters than those who did not. Accordingly, before discussing the subject of criterion validity, it seems appropriate at this point to turn to the subject of the test preparation and, then, to the method of scoring.

C. Preparation of the Test

The standard required for test preparation is that the test makers exercise reasonable competence, so that it can reasonably be said that the matters measured by the test are the matters identified as appropriate for measurement by the job analysis. Guardians IV, supra, at 95. Here, there can be no

question that AIR and Dr. Fleishman exercised reasonable competence in the design of their factorially pure tests to measure the constructs identified by the job analysis as necessary to the job. Whatever disagreement there may be within the test maker's profession as to whether Dr. Fleishman's method of test construction will ultimately prove the best one, there can be no question that it represents one reasonable approach to the subject, carried out, at least in instances which preceded and followed the actual preparation of Exam 3040, according to a high level of professional competence.

The problem which arises in this case, however, is that Dr. Fleishman and his subordinates were subjected to marked interference in the process of professional job preparation by representatives of the Fire Department and of the

City's Department of Personnel. As we said in Guardians IV, supra, 630 F.2d at 96:

> "Of course, the law should not be designed to subsidize specialists. But employment testing is a task of sufficient difficulty to suggest that an employer dispenses with expert assistance at his peril."

Here, the City, under Judge Weinfeld's mandate, employed experts to assist them, but then took the dangerous course of dispensing with the experts' advice. The resulting mixture of professionalism and the kind of subjective judgment of supervisors which condemned an earlier generation of employment tests, see Guardians IV, supra, 630 F.2d at 88-89, resulted in an inadequate job of test preparation.

Most notable of the defects in test preparation is the substitution of the series of so-called face valid tests for the factorially pure tests selected

by Dr. Fleishman after extensive analysis and testing. As noted, the agility test and dummy test first appeared between the December 1973 pre-test and the February 1974 proposed test battery not because of any consideration having to do with job-relatedness, but apparently in an effort to accommodate the demands made on the City by other disadvantaged groups involved in the Vulcan litigation. While the motivation behind this substitution was certainly laudable, the high purpose is no excuse for a marked relaxation of professional standards. In the case of the agility test, the effect of substituting it for the cable test earlier conceived of as the appropriate one for measuring the ability of gross body coordination was to introduce marked overlapping and distortion in the rank ordering of the abilities purportedly identi-

fied as necessary for the job of fire-fighting. While an effort at factor analysis of the agility test appears to have been made, the results it produced -- namely, low correlation between the test and gross body coordination -- dictated against the substitution rather than in favor of it.

The crucial eight-foot wall which entered the test battery by means of the agility test was justified not in terms of the PAA technique otherwise used to analyze the job's requirements, but rather in terms of an altogether ad hoc and incomplete inquiry into job content -- whether New York City firemen ever have occasion to climb eight-foot walls in the course of the job. As noted earlier, this inquiry failed to determine the importance, frequency, or criticality of the job behavior or even to define it

with sufficient precision to make it possible to judge whether the scaling techniques required were those appropriately measured by the obstacle course wall.

Similarly, the dummy carry, derived from the February 1974 height study also carried out as an outgrowth of the Vulcan litigation, introduced further distortions of ranking and overlapping of tests into the factorially pure battery devised by Fleishman. As the evidence at trial established, the carry in fact tested in Physical Exam 3040 tests for abilities different than those that are actually needed in rescue operations and, to the extent it does test for abilities needed in one form of firemen's carry, those abilities are ones learned by training rather than received as part of one's natural endowment.

Equally disquieting in terms of the professionalism of the test preparation is the elimination of AIR's balance and twist-and-touch tests as measuring abilities needed by ballet dancers as opposed to firefighters. Again, the failure of the professional test maker to turn the discussion from the dangerous realm of job images to the neutral realm of job requirements must be criticized both in terms of the overall goals of the test preparation and in concrete terms of test preparation. The insertion of what was deemed to be a test of extent flexibility (the window-ladder-window test) into the agility run meant that no accurate test of extent flexibility would be recorded at all unless the candidate mastered the eight-foot wall which, whatever else it did, did not measure for extent flexibility. The substitution of

the ledge walk for the balance test (to be performed with eyes closed to insure that gross body equilibrium was measured rather than balance achieved through visual contacts) again compromised the factorial purity of AIR's battery. It also introduced an observable behavior into the test without any factual inquiry as to the existence and importance in firefighting of having to side step at high speed forward and back over a narrow ledge.[20]

---

[20] Defendants seek to excuse shortcomings in the job analysis and test preparation by reference to Spurlock v. United Airlines, 475 F.2d 216, 219 (10th Cir. 1972), which determined that, because of the economic and human risks of hiring unqualified flight officers to pilot what was described as a $20 million aircraft carrying as many as 300 passengers per flight, "the employer bears a correspondingly lighter burden to show that employment criteria are job related." Whatever the merits of applying such a rule to the hiring of firefighters (see Vulcan Society v. Fire

(Footnote continued)

Of similar significance in evaluating AIR's test preparation is its "tightening" of the test requirements in response not to any requirements of the job, but rather in response to the "subjective judgments of supervisors," Vulcan Society v. Civil Service Commission, supra, 490 F.2d at 396-98. While there is certainly every reason for a profes-

---

(Footnote 20 continued from previous page) Department, 505 F. Supp. 955, 965 (S.D.N.Y. 1981), in which Judge Sofaer declined to impose a requirement of a high school diploma on firefighters in the face of an argument that it was required by the Spurlock rule), it is clear that its application cannot save a job analysis and test preparation which fail not simply as a matter of degree, but in the overall logic and consistency of the approaches employed. Thus, even accepting the rule of the Spurlock case as appropriately employed to New York City firefighters because of the economic and human risks involved in their work, the inconsistencies in the test preparation and in defendants' attempts at validation prevent me from validating the test on that ground.

sional to listen to the criticism of job incumbents on the lack of relation of the test to the job's requirements,[21] the reason for listening to the criticism is to make certain that the test in fact tests for what the job needs. Here, a beneficial result of the criticism of the one-mile run was that AIR recognized an error in its test preparation, namely, that the test would not be performed on the 31-lap track on which the norm scores had been based. However, instead of then re-doing the pre-test or re-calculating it to eliminate the time attributable to the test takers having to run more corners than would be present in the actual

---

[21] There is considerably less justification for relying on the comments of a non-job incumbent, such as a representative of the Department of Personnel, as to what it takes to be a firefighter.

test, thereby assuring that the norms would be based on a sample of the applicant group as originally intended, AIR accepted from the Department of Personnel and the Fire Department a set of norms having little if any rational connection with what it was reasonable to expect either of applicants or of incumbent firemen. Thus, instead of using the norms developed on a group of probationary firemen and firemen in training, reasonably thought to approximate the class of applicants as it was then anticipated it would be, AIR used figures for the mile run developed by Thomas Cureton in studies conducted in the earlier '40's on men in physical training for war. The combined effect of the effort to tighten up the mile run requirements was to produce a scoring table for the mile run which not only could not be passed by any

of the incumbent firemen who took the December 1973 tryout, but also would have been failed by half of the Cureton sample.

Similar tightening of the scoring was based on subjective criticism by Fire Department personnel and representatives of the Department of Personnel of the cut-off scores for the broad jump, the hand grip, and the dummy carry. This overall tendency to defer to the pressures of employers to get "the best men" for the job, without consideration whether such a requirement unnecessarily excludes women or other disadvantaged groups, requires a conclusion that the test preparation of Exam 3040 was wanting in professional competence precisely because it failed to take reasonable steps to exclude engrained discrimination of the sort that Title VII was designed

to eliminate. Griggs v. Duke Power Co., supra at 429-30.

D. Test Scoring

As already noted, the City did not use the results of the exam simply to determine who possessed the minimum qualifications necessary for the job, as it had on earlier physical examinations administered on a pass/fail basis in 1962, 1968, and 1971 (Exams 9606, 7060 and 0159). Instead, it used the results of the test (after averaging the raw scores on both the written exam and the physical) to compile the rank-ordering of all applicants on a scale of 1,000 (100.0%) with a cut-off score fixed at 700 (70.0%). The use of a percentage scale with ranks calculated to the nearest tenth of a percent and the use of the 70th percentile as the passing score appears to have been decided upon simply

because such a rank and cut-off method was customary in competitive civil service testing and, indeed, had been the practice on earlier competitive firefighters' exams.

As was stated in Guardians IV, supra, 630 F.2d at 100:

> "The Guidelines provide that rank-ordering should be used only if it can be shown that 'a higher score . . . is likely to result in better job performance.' Guidelines § 14(C)(9). This requirement is reasonable and consistent with Title VII's provision that the 'results' of a test may not be 'used to discriminate.' 42 U.S.C. § 2000e-2(h). If test scores do not vary directly with job performance, ranking the candidates on the basis of their scores will not select better employees . . . . The frequency with which such one-point differentials are used for important decisions in our society, both in academic assessment and civil service employment, should not obscure their equally frequent lack of demonstrated significance."

At first glance it might seem that, since no women passed the physical exam, the only subject that need concern

us is the validity of the cut-off score. Realistically, however, since over 17,000 persons took the physical exam and since fewer than 3,000 persons have been appointed to date from the all-male eligibility list on which over 8,000 names appeared, the effective cut-off score is as a practical matter considerably over the 70th percentile. Accordingly, it is the strict rank ordering of Exam 3040, pursuant to which every one of the thousands of successful candidates in the 299 ranks between 100.0 and 70.1 must be exhausted before the candidates in the first passing rank are called, that must be considered to determine whether it is predictive of job performance.[22]

---

[22] Significantly, the one woman firefighter who testified in the trial of this case -- a member of the Chicago fire department -- related that she had become a firefighter, after securing a passing score of just above 70

(Footnote continued)

Based on the evidence introduced at the hearing, I conclude that here, as in Guardians IV, supra, 630 F.2d at 101, "the defects . . . in the job analysis and the test construction are substantial enough to preclude an inference that passing scores will correlate with job performance closely enough to justify rank-ordered selections" of the type which resulted from Exam 3040.

First, as has been noted by the EEOC, it is "easier" to make the inference between higher scores and better job performance, "[t]he more closely and

---

(Footnote 22 continued from previous page) and, as a result, being rank ordered to the bottom of Chicago's eligibility list, solely as a result of the happenstance of a firefighter's strike which resulted in the call-up of the entire eligibility list. She has been serving now in firefighting operations in Chicago for close to two years, by all accounts, with distinction.

completely the selection procedure approximates the important work behaviors." EEOC, Uniform Employee Selection Guidelines: Interpretation and Clarification (Questions and Answers) Q. 62 (1979). Here, the lack of systematic study of the relationship between work behaviors observed in the PAA and PDA and the abilities tested for in the test battery makes it apparent that the fine gradations of the rank order employed are hardly justified by the job analysis. Moreover, an examination of the PAA and PDA instruments themselves makes clear that for the most part all that AIR hoped to establish were gross distinctions in the degrees of abilities and physical demands, respectively. For example, in the PDA, ratings for balancing are in terms of "no importance," "important," "high importance"; weight-lifting ability is analyzed in

terms of five categories of weight. In the PAA all abilities are rated on a scale of 1 to 7. While analyses of this sort might be used to justify a limited series of grades (within which the candidates would be selected at random), they cannot be used to justify the extraordinary pretense at precision implicit in the ranking established by Exam 3040.

In terms of test preparation, the compromise in the weighting of the subtests, resulting from the substitution of the multiple-ability tests for the tests selected originally by AIR and the "tightening" of the test scores in response to subjective judgments of the employer that firefighters ought to be able to run the mile faster, jump farther, and have a firmer hand grip than the experts concluded, likewise makes it impossible to say with any reasonable

measure of confidence that the rank order resulting from the test is an accurate predictor of job performance.

The City responds to the attack on its rank ordering not with any justification drawn from its job analysis or based on its test preparation, but rather by an assertion that every increment in the abilities tested for in its physical exam necessarily represents a better performance as a firefighter. This conclusion was presented as "obvious" by two witnesses, one an industrial psychologist, the other a physiologist who was also a volunteer fireman, both of whom testified that it is in the "nature" of firefighting that the stronger the firefighter the better. This proposition was, however, hotly contested by plaintiff's witnesses, including an industrial psychologist with extensive acquaintance

with firefighting and an exercise physiologist, among others, who pointed out that few jobs making large physical demands, least of all firefighting, are properly performed at maximum speed or at the limits of one's strength or endurance. As these witnesses testified, what must be identified are not those who are strongest or fastest but, instead, those who, with the benefit of training in pacing or because of their native capacities of endurance, can perform the punishing tasks of firefighting as they are actually required to be performed. According to these witnesses, firefighting takes its toll, not as a result of failures of maximum strength or speed, even at critical moments, but rather through the physical demands extending over long periods of time which necessitate paced performance at less-than-maximum levels.

This explanation makes sense of a number of otherwise puzzling features of fire-fighting, namely, the ability of firemen to continue to do their work competently over an entire working career and the results of physical testing of incumbent male firemen which show that on a variety of measures of maximum physical capacity firemen rate no better than the average American male. It also makes sense in terms of the recognized dangers of fire-fighting and their unpredictability which, as numerous witnesses testified, make hazardous in the extreme performance at top speed or at the limits of strength or capacity.[23] Not only does maximal

---

[23] An illustration of this point is contained in a comment by an AIR observer who recorded his on-the-job observations while accompanying firemen on several runs (D.X. CC, pp. 36-37):

"I observed at a tenement fire in Brownsville a situation where the

(Footnote continued)

performance compromise the firefighter's ability to pace performance over the long periods of time during which physical demands are being made on the body, in addition, the very unpredictability of fire and the instability of burning structures call upon qualities of foresight, endurance, and pacing not examined by tests of maximum physical strength.[24]

(Footnote 23 continued from previous page)
Lieutenant gave a probie fireman immediate feedback about his aggressiveness. For whatever reason, this probie apparently was moving upon the firefront[;] however, he was doing this at a rate which the Lieutenant considered too fast just for safety, namely, he wasn't checking and making sure that the fire didn't break out behind him."

[24] One illustration of this distinction is the contrast between the ledge balance test, calling upon the test taker to race at top speed sideways along a balance beam said to represent a ledge facing a wall, and balancing as described in the examples furnished in the PDA and in the testimony of trial witnesses. What is required is
(Footnote continued)

In all events, even the industrial psychologist who testified that every increment of strength predicted a better fireman did not claim that his own test instrument could grade firefighters beyond the degrees of ability measured on a scale of excellent, good, average, fair, and poor (P.X. 251, Table 5). In short, neither the job analysis instrument, the test instrument, nor the validation instruments appear able to perform their tasks with the precision necessary to justify the rank ordering used here.

This conclusion does not necessarily mean that a future exam must be administered on a pass/fail basis. It does suggest, however, that some larger

---

(Footnote 24 continued from previous page) the ability to get across an exposed floor joist, for example, safely within a reasonable time, not top speed.

use of random selection within fewer, more rationally grounded ranks will have to be substituted for the present system, unless finer test instruments can be found.

E. The Effort to Criterion Validate Exam 3040

There remains for consideration whether Exam 3040 can be criterion validated by a study of firefighters' task performances conducted at the University of Maryland for the National Fire Data Center subsequent to the administration of Exam 3040.

The City's "borrowing" of this study of a group of firefighters from the Washington, D.C. area was offered late in the trial in an effort to obtain predictive validation for Exam 3040 pursuant to Guidelines § 7 which provides:

A. Validity studies not conducted by the user. Users may, under certain circumstances, support the use of selection procedures by validity studies conducted by other users or conducted by test publishers or distributors and described in test manuals. While publishers of selection procedures have a professional obligation to provide evidence of validity which meets generally accepted professional standards (see section 5C above), users are cautioned that they are responsible for compliance with these guidelines. Accordingly, users seeking to obtain selection procedures from publishers and distributors should be careful to determine that, in the event the user becomes subject to the validity requirements of these guidelines, the necessary information to support validity has been determined and will be made available to the user.

B. Use of criterion-related validity evidence from other sources. Criterion-related validity studies conducted by one test user, or described in test manuals and the professional literature, will be considered acceptable for use by another user when the following requirements are met:

(1) Validity evidence. Evidence from the available studies meeting the standards of section 14B below clearly demonstrates that the selection procedure is valid;

(2) Job similarity. The incumbents in the user's job and the incumbents in the job or group of jobs on which the validity study was conducted perform substantially the same major work behaviors, as shown by appropriate job analyses both on the job or group of jobs on which the validity study was performed and on the job for which the selection procedure is to be used; and

(3) Fairness evidence. The studies include a study of test fairness for each race, sex, and ethnic group which constitutes a significant factor in the borrowing user's relevant labor market for the job or jobs in question. If the studies under consideration satisfy (1) and (2) above but do not contain an investigation of test fairness, and it is not technically feasible for the borrowing user to conduct an internal study of test fairness, the borrowing user may utilize the study until studies conducted elsewhere meeting the requirements of these guidelines show test unfairness, or until such time as it becomes technically feasible to conduct an internal study of test fairness and the results of that study can be acted upon. Users obtaining selection procedures from publishers should consider, as one factor in the

decision to purchase a particular selection procedure, the availability of evidence concerning test fairness.

The first matter which bears noting is that defendants have not offered any evidence of the "test fairness" of Exam 3040 for members of the female sex as required by Guideline § 7B(3), supra. Moreover, the only evidence offered by defendants to support an argument that it is not technically feasible for New York City to conduct such a study of the fairness of the test is that New York City has at present no female firefighters whose performance on the job could be measured against their performance on Exam 3040. The answer to this argument is that the criterion used to measure the accuracy of the predictors in the Maryland study was not performance in actual firefighting tasks by job incumbents. Instead, the criterion was per-

formance on a series of so-called job sample tasks, requiring neither instruction nor training. New York City's lack of female job incumbents was, in other words, no barrier at all to the performance of the series of Maryland criteria tasks by a representative sample of male and female firefighter applicants in order to determine the fairness of applying those criterion tasks to measure the accuracy of Exam 3040 as a predictor of job performance. However, quite apart from that fundamental objection, the University of Maryland study does not purport to validate the tests administered in Exam 3040 or anything like them.

What the Maryland study does is suggest both a laboratory and a field test to predict performance ability of firefighters. The laboratory test consists of a determination of percent of

body fat, lean body weight, maximal heart rate, score on a treadmill test, grip strength, performance of sit-ups and long jump, and submaximal oxygen pulse. Of this list, only the long jump and the grip strength tests bear any resemblance to the subtests administered as part of Exam 3040, and in fact both of these differ in either their method of administration or scoring from the tests administered in New York. The field test consists of a determination of percent fat, lean weight, score on a step test, push-ups, sit-ups, and grip strength. In this group, only grip strength bears resemblance to a subtest administered as part of the New York City exam.

Defendants do not, in fact, suggest that Exam 3040 is comparable to either of the two recommended Maryland tests. Instead, by picking and choosing

from among the subtests actually recommended in Maryland, from among a number of other predictor tests considered and rejected in the Maryland study, and, finally, from among the performance measures or tasks used in Maryland in an effort to validate the predictor tests recommended, defendants came up with a group of five "matching" tests, the validation of which they claim validates Exam 3040.[25] Even this extraordinary selec-

---

[25] The dangers of this kind of picking and choosing are manifest. For one thing, such a process severely undercuts one rationale behind permitting the borrowing of another's test, namely, that the test has been actually chosen and either sold (by a test manufacturer or test developer) or at least adopted by another test user willing to stand behind its results. Indeed, the borrower in such a situation as that presented here is not, in fact, relying on any effort on the part of a test lender to criterion validate a specific test battery, but is, instead, simply asserting that individual tests, e.g., hand grip or broad jump, have been shown by studies

(Footnote continued)

tion process, in other words, could not secure a match for all seven of the subtests employed in Exam 3040; and, in fact, the five performance measures selected to match the subtests of Exam 3040 do so only approximately.

In addition to the Maryland long jump and hand grip, the defendants chose two predictor measures which were considered, but not recommended, as part of the Maryland test batteries, namely, a balance beam test and a 12-minute run, as matching the New York City balance test and mile run, respectively. Further, a measure used in the Maryland study in an effort to validate the Maryland tests, namely, a dummy drag or carry, was chosen

---

(Footnote 25 continued from previous page)
to be good predictors of job performance without consideration of their job-relatedness, representativeness or fairness.

by the defendants as a match for the New York City dummy carry which was in New York, of course, a predictive measure.[26] The distinctions between each of these three tests and the New York City tests are considerable.

The 12-minute run used in Maryland measures the distance run by a candidate in 12 minutes and was almost universally acknowledged by the experts at trial to be superior to the mile run as a test of stamina since it assists the runner in finding his or her own pace, thereby more accurately measuring aero-

---

[26] This readiness to ignore what the authors of the Maryland study thought they were doing, by converting one of the Maryland criterion measures into a predictive measure, in order to establish the validity of the New York City exam, makes clear that defendants are not, in fact, relying on the Maryland study for anything like what it purported to conclude about the validity of firefighter's tests.

bic, as opposed to anaerobic, capacity. The balance test employed in Maryland, in contrast to the New York City ledge walk, involved walking forward over a balance beam with no time limit or emphasis on speed. Finally, the Maryland dummy drag consisted of dragging a dummy down five flights of stairs, as opposed to the test in New York involving a carry of the dummy first up and then down a flight of stairs. Clearly, the substantial differences between these three tests and their alleged counterparts in New York preclude validation of the New York City tests by such validation as they were given in Maryland.[27] Since all of the criterion validation measures offered by defendants

[27] It bears repeating that the Maryland dummy drag was never, in any sense, validated in Maryland since it was in fact one of the criterion measures by which other predictive measures were validated.

as established by the Maryland study include at least one of these markedly dissimilar tests on the predictor side, it can hardly be said, laying all its other faults aside, that defendants' efforts to borrow the Maryland study to validate the New York test have, to any degree, succeeded.

However, there are even more substantial problems with defendants' criterion validation effort. The method by which these Maryland tests were validated was by a study of the correlations between the results of the performance of these tests by a group of firefighters from the Washington, D.C. area and the scores they obtained in performing five tasks said to replicate actual firefighting behavior. These tasks involved (1) the vertical raising of a 35-foot ladder; (2) the carrying of 73 lbs. of

standpipe hose up five flights of stairs; (3) the pulling of 52 lbs. of hose to the fifth floor window by means of a line; (4) the dragging of a 117-lb. dummy down five flights of stairs; and (5) striking a railroad tie with 30 swings of an eight-lb. sledge hammer.

In the Maryland study the researchers measured the subjects' heart rate and time to complete these five tasks in order to determine which subjects had performed the work fastest with the greatest reserve ability to continue working. The ability of the selected tests to predict high performance on the criterion tasks was seen as validating the predictors. In borrowing the Maryland study for validation purposes, however, defendants inexplicably eliminated the average heart rate as a factor in rating task performance and sought to

validate the tests said to be comparable to some of those used in New York City based solely on the speed with which the subjects completed the five criterion tasks. The faster one accomplished the five tasks, all to be performed seriatim as expeditiously as possible, the more successful one was considered as a fireman for the purpose of validating the New York City predictive tests.

The problem with this method of analysis is clear. Even if one could agree based on an adequate job analysis that the five job sample tasks simulate actual job behaviors, there is here obviously an assumption about the appropriate way to measure optimum performance of the tasks which has no basis in the evidence or common sense, that is, that the most successful firefighter will be the one who performs these five tasks in the

shortest time. Moreover, when one realizes that the characteristic manner in which the dummy test was performed in the Maryland test was to drag the dummy down five flights of stairs at top speed, one begins to question whether the job tasks in fact do simulate actual job behavior. Moreover, giving speed the emphasis it received in the Maryland study, even taking into account heart rate, appears highly questionable. This seems obvious, once it is realized that the mean time to complete the five simulated tasks was 7.03 minutes, during which the firefighters' average heart rate was measured at a value representing 97 percent of the maximum heart rate. This heart rate was taken by the Maryland researchers to suggest that firefighters, while performing their tasks, work at their maximum aerobic capacity. However, as the study

itself paradoxically noted as a result of comparing these demands with the general fitness level of the firefighters studied, "it can be concluded that the aerobic capacity of the firefighter is not, on the average, adequate to complete typical firefighting tasks at the pace observed in this study." The study then concluded that the firemen studied were unfit for their jobs. If the firemen used to validate the tests were unfit for the job, it can well be asked how their performance could be used to validate tests administered in other jurisdictions. A much more obvious explanation for the apparent paradox is that the pace at which the tasks were performed bore no relation to reality and that the validity study does not pass the technical standard for validity studies in Guidelines § 14B(3) that "[w]hatever criteria are

used should represent important or critical work behavior(s)." In all events, nothing in the Maryland study shows that performance of the five tasks listed above under actual fighting conditions, assuming them to be important or critical work behaviors, is to be rated simply according to the speed with which they can be performed seriatim. Accordingly, I can find no basis in the Maryland study for determining that Exam 3040 is criterion valid -- that is, that it accurately predicts the abilities of the members of the plaintiff class who took the physical exam to be New York City firefighters.

II

Relief

In considering the relief to which plaintiff is entitled, it is appropriate "to distinguish between those aspects . . . designed to assure compli-

ance with Title VII and those aspects that provide affirmative relief as a remedy for past discrimination." Guardians IV, supra, 630 F.2d at 108. Turning first to the relief necessary to assure that Title VII's requirements will be complied with, "[c]ompliance involves restricting the use of an invalid exam, specifying procedures and standards for a new valid selection procedure, and authorizing interim hiring that does not have a disparate impact." Id.

In order to insure that an invalid exam is not improperly used to discriminate in the future, it appears appropriate in this case, as it was in Vulcan Society v. Civil Service Commission, supra, 360 F. Supp. at 1278, to enjoin defendants "from making any further appointments based upon its results, without prejudice, however, to any appli-

cation by the parties, upon a showing of compelling necessity, for interim relief which would permit appointments from the eligible list upon an equitable basis pending the prompt promulgation and administration of a new examination free from . . . taint." As Judge Weinfeld noted in the Vulcan Society case, "to freeze all appointments [to the Fire Department] may present a hazardous situation to the citizens of the community."

It is also appropriate, however, to direct defendants to commence forthwith preparation of new and valid selection procedures, validated in general accord with the Guidelines and the APA Standards and which have the least adverse impact on women. Id.

With respect to interim hiring provisions, such as are sought here:

> "[s]ince interim hiring provisions, where needed to satisfy immediate personnel requirements, are to be used prior to development and approval of a valid selection procedure, such provisions cannot meet Title VII standards by demonstrated job relatedness. Therefore, one appropriate way to assure Title VII compliance on an interim basis is to avoid disparate . . . impact. This means selecting from among adequately qualified applicants either on a random basis, see, e.g., Association Against Discrimination, supra, 594 F.2d [306] at 313 n.19 [2d Cir.], or according to some appropriately noncompensatory ratio, see, e.g., Kirkland, supra, 520 F.2d [420] at 429-30 [2nd Cir.]; Vulcan Society, supra, 490 F.2d at 398-99, normally reflecting the minority ratio of the applicant pool or the relevant work force." Id. at 109.

In this case since it cannot be said that Exam 3040 was sufficient to determine which among the plaintiff class were "adequately qualified," it will first be necessary to determine which, if any, of the plaintiff class are in that category, either by a procedure to be agreed upon by the parties or, in the

absence of agreement, by one established by the Court.

With regard to the number of positions which will be offered to qualified women applicants determined by this means, the percentage of women in "the relevant work force" does not, as already noted, appear an adequate measure of non-discriminatory hiring. As defendants point out, it is unreasonable to suppose that the relevant work force can be determined without further evidence concerning the physical abilities of those women in the work force to do the physical work required of firefighters. Instead, it appears more appropriate in this case to resort to the actual "applicant pool," id., to determine the appropriate measure of non-discriminatory hiring, since the applicant pool represents those women in the work force who,

like their male counterparts, believed themselves to be qualified for the job.

At the same time, care must be taken to avoid giving continued force to the deterrent effect of the City's discriminatory physical test as a result of the publicity and familiarization program. While it is difficult to calculate with any precision the deterrent effect of the physical exam on those who took and passed the written exam, one reasonable measure results from comparing the approximately 77% fall-off in interest among women who took and passed the written exam with the approximately 26% fall-off in interest among men in the same category. Assuming that the sex-neutral exam would have been followed by an equal decline (26%) in interest between both groups, one would have expected a group of approximately 288 women to have pre-

sented itself to take the physical exam. Finally, it seems reasonable to assume that, given a sex-neutral exam and sex-neutral call-up rates, the ratio of persons called up to total applicants would be the same for women and men. Of the total of 16,925 men presenting themselves to take the physical exam, 2,666, or approximately 16%, have been called up to date. Applying the same percentage figure to the 288 women who one would have expected to have taken the physical exam but for its deterrent effect results in a figure of close to 45 women who one would expect to have been called up under a sex-neutral exam. This figure compares favorably with the figure of 50 agreed upon by the parties as reflecting the number of places appropriate to reserve for women, should the exam be found to be discriminatory. In the absence of some

other proposal, the most reasonable, non-discriminatory measure to provide for interim hiring appears to be 45 places.

Plaintiff recognizes, however, given the small size of the plaintiff class to begin with, the deterrent effect of the exam and the delay which has occurred as a result of the preparation of this case for trial, that it may well be that there are not available 45 women in the class as originally defined who continue to be interested in the position of firefighter. Nevertheless, plaintiff argues that she is entitled to a decree directing that defendants affirmatively recruit and appoint women for these positions either because defendants' discrimination was intentional or because there has been demonstrated here a pattern of significant and long-standing discrimination warranting such relief.

I find no basis in the record for a determination that defendants' discrimination against women has been intentional. Nor do I find that the evidence is sufficient to warrant a conclusion that there has been "a demonstrated pattern of significant prior discrimination." Guardians IV, supra, 630 F.2d at 112.

With regard to intentional discrimination, apart from a few isolated remarks demeaning to women by referring to them as a group in circumstances in which it was appropriate to deal with them as individuals, the only evidence relied on is the continued use by defendants of the eligibility list established pursuant to Exam 0159 which was closed to women after 1972 when Title VII was amended to include municipal employers.

However, as noted in Guardians IV, supra at 112, in rejecting a similar argument:

> "[p]ersistent use of an exam with disparate . . . effects would support an inference of intentional discrimination if proper test construction were not even attempted." (Emphasis added)

Here, as in Guardians IV, proper test construction was at least attempted: "The City's . . . officials made extensive efforts to understand and apply the Guidelines and develop a test they hoped would have the requisite validity." Here, unlike Guardians IV, at least one warning concerning the adverse impact of the test on women was, in fact, heeded and led to the substitution of a test conceded to have a less adverse effect on women. That defendants proceeded in what Guardians IV referred to as "a naive self-confidence" that this was the only aspect of the exam which adversely af-

fected women is not a sufficient basis for a finding of intentional wrong-doing.

Nor is the evidence sufficient to establish a demonstrated pattern of significant prior discrimination. To be sure, of the six examinations administered between 1960 and 1978, which are responsible for the present composition of the fire department, none was open to women. No women are firefighters at present. Moreover, women constitute 44% of the labor force between the ages of 18 and 29 in the relevant geographic area. However, as indicated above, there is at present insufficient evidence from which to conclude that a substantial part of the total female work force possesses the requisite qualifications necessary to perform the job of firefighter. In the absence of proof that any substantial part of the female work force possesses

the requisite qualifications for the job, there has been no showing of the kind of "flagrant disparity shown in prior cases where long-term hiring quotas were in issue." Guardians IV, supra, 630 F.2d at 113.

Accordingly, defendants are hereby enjoined from further use of the eligibility list compiled pursuant to Exam 3040, except upon a showing of compelling necessity. Notices of appointment shall be sent to up to 45 of those women who in 1977 and 1978 applied to become firefighters and who are found to be qualified for appointment and willing to be appointed.[28] In the event that more than 45 are found to be qualified, then notices of appointment shall be sent

---

[28] In the event that more than 45 members of the class are found qualified and willing to accept appointment, selection shall be by lot.

to 45 of those qualifying, selected by lot.[29] Qualification shall be determined pursuant to procedures agreed upon by the parties. In the event the parties are unable expeditiously to agree upon such procedures, then the procedures shall be, on notice to the parties, fixed by the Court.

The Clerk is directed to notify the attorneys for the parties and intervenors of the entry of this Memorandum Decision and to mail a copy to each.

Settle Order on Notice.

Dated: Brooklyn, New York
March 4, 1982

/s/ Charles P. Sifton
United States District
Judge

---

[29] If more men are called up between the date hereof and the date of the list's expiration, it will be appropriate to increase the number of places available to be filled by qualified women, correspondingly.

UNITED STATES COURT OF APPEALS
FOR THE
SECOND CIRCUIT

At a stated Term of the United States Court of Appeals for the Second Circuit, held at the United States Courthouse in the City of New York, on the seventeenth day of February one thousand nine hundred and eighty-seven.

Present: HON. WILFRED FEINBERG,
Chief Judge,

HON. JON O. NEWMAN,

HON. AMALYA L. KEARSE,

Circuit Judges,

BRENDA BERKMAN, on behalf of herself and a class consisting of all similarly-situated women,

Plaintiff-Appellee,
Cross-Appellant,

- v. -

THE CITY OF NEW YORK; EDWARD I. KOCH, individually and as Mayor of the City of New York; NEW YORK CITY FIRE DEPARTMENT; AUGUSTUS BEEKMAN, individually and as Fire Commissioner of the City of New York; NEW YORK CITY DEPARTMENT OF PERSONNEL; MICHAEL NADEL, individually and as Director of Personnel of the City of New York; CIVIL SERVICE COMMISSION OF THE CITY OF NEW YORK,

Defendants-Appellants,
Cross-Appellees,

Nos: 86-7157,
86-7159,
86-7167,
86-7201

-and-

UNIFORMED FIREFIGHTERS ASSOCIATION, LOCAL 94, FIREFIGHTERS ELIGIBLES ASSOCIATION, LIST NO. 1162, INC.,

Defendants-Intervenors,
Appellants-Cross-Appellees,

-and-

JAMES T. AHRENS,
Defendant-Intervenor.

Appeal from the United States District Court for the Eastern District of New York

This cause came on to be heard on the transcript of record from the United States District Court for the Eastern District of New York, and was argued by counsel.

ON CONSIDERATION WHEREOF, it is now hereby ordered, adjudged and decreed that the Order of said District Court be and it hereby is affirmed in part, reversed in part and remanded to the said district court for further proceedings in accordance with the opinion of this court.

Elaine B. Goldsmith,
Clerk

By: s/ Edward J. Guardaro
Edward J. Guardaro,
Deputy Clerk

UNITED STATES COURT OF APPEALS
FOR THE [SECOND] CIRCUIT

At a stated Term of the United States Court of Appeals, in and for the Second Circuit, held at the United States Courthouse, in the City of New York, on the 14th day of April one thousand nine hundred and eighty-seven.

BRENDA BERKMAN, on behalf of herself and a class consisting of all similarly-situated women,

Plaintiff-Appellee,
Cross-Appellant,

v.

THE CITY OF NEW YORK; EDWARD I. KOCH, individually and as Mayor of the City of New York; NEW YORK CITY FIRE DEPARTMENT; AUGUSTUS BEEKMAN, individually and as Fire Commissioner of the City of New York; NEW YORK CITY DEPARTMENT OF PERSONNEL; MICHAEL NADEL, individually and as Director of Personnel of the City of New York; THOMAS ROCHE, individually and as former Director of Personnel of the City of New York,

Nos: 86-7157
86-7159
86-7167
86-7201

Defendants-Appellants,
Cross-Appellees,

-and-

UNIFORMED FIREFIGHTERS ASSOCIATION, LOCAL 94, FIREFIGHTERS ELIGIBLES ASSOCIATION, LIST NO. 1162, INC.,

Defendants-Intervenors,
Appellants-Cross-Appellees,

-and-

JAMES T. AHRENS,

Defendant-Intervenor.

A petition for rehearing having been filed herein by counsel for the plaintiff-appellee, cross-appellant, Brenda Berkman, on behalf of herself and a class consisting of all similarly situated women,

Upon consideration thereof, it [is]

ORDERED that said petition be and it hereby is DENIED.

Elaine B. Goldsmith,
Clerk

By: /s/ Marcia C. Mahan
Deputy Clerk

UNITED STATES DISTRICT COURT
EASTERN DISTRICT OF NEW YORK

- - - - - - - - - - - - - - - - - -x

BRENDA BERKMAN, :

Plaintiff, : 79 C 1813

- against - :

THE CITY OF NEW YORK, :
et alia,
:
Defendant.
- - - - - - - - - - - - - - - - - -x

United States Courthouse
Brooklyn, New York

June 13, 1986
2:00 p.m.

B E F O R E:

HONORABLE CHARLES P. SIFTON, U.S.D.J.

A P P E A R A N C E S:

DEBOVISE [SIC] & PLIMPTON
BY: JONATHAN E. RICHMAN, ESQ.
LAWRENCE M. MARTIN, ESQ.

BY: FREDERICK A. O. SCHWARZ, JR.
CORPORATION COUNSEL
NORMA KERLIN
ELIZABETH DALE KENDRICK

DAVID V. KIRBY
Assistant United States Attorney

Also present: Vicki Been, Esq.

THE COURT: Who is here representing the plaintiff?

MR. MARTIN: Lawrence M. Martin, Your Honor, for the Plaintiff Debevoise and Plimpton.

THE COURT: Who is here representing the defendant?

MS. KERLIN: Norma Kerlin.

THE COURT: Are there any other appearances?

THE COURT: I have asked a representative of the U.S. Attorneys Office to be present during these proceedings for reasons which will become apparent.

Is there anything further that either side wants to add to the papers that have been submitted?

MS. KERLIN: No, Your Honor.

THE COURT: Ms. Kerlin, since you have presented some objections, is there anything that you want to say.

MS. KERLIN: No, Your Honor.

THE COURT: Anything further the plaintiff wishes to add?

MR. RICHMAN: I would be happy to address any of the interests [sic] that interests [sic] you. Barring that the answer is, no.

THE COURT: All right. Well, I am prepared, given the short time before expiration of the temporary restraining order that is currently in effect, to render a decision with regard to the objections that have been filed to the Magistrate's report and recommendation.

Please have a seat and I'll give you the decision I have reached.

As you are aware, this matter was referred to the Magistrate to prepare a report and recommendation with regard to the application made by Firefighter Katrina Cannon seeking to enjoin, pending the final disposition of her complaint, and [sic] the transfer that was proposed to be made of her from an engine company to which she is now assigned, Engine Company 274, the 52nd Battalion, 14th Division in Queens.

The Magistrate has conducted hearings and an evidentiary record has been prepared.

On the basis of that record he has filed a report and recommendation that the preliminary injunctive relief be granted. The defendants have timely served and filed written objections to the Magistrate's report.

I was alerted to the fact that there would be a very short time for review of the record with respect to any objections made already. Accordingly, I have already started a de novo review or started a de novo review by reading through this voluminous record, and now in the light of the objections that have been made I have concluded the de novo review of those portions of the recommendation, to which objections to [sic] have been made.

Having made that review I have determined to accept the findings and recommendations that have been made by the Magistrate with one exception which I will discuss in a moment.

Based on those findings and pursuant to those recommendations, I am going to sign and file today a prelimi-

nary injunction granting Firefighter Cannon the relief which she seeks.

The defendants have made four objections to the Magistrate's report. First, they contend that the Magistrate's conclusion, that the firefighter will be irreparablely [sic] harmed if the injunctive relief is not granted, is contrary to the law and to the evidence; secondly, the defendants contend that the Court is without jurisdiction over the matter because of the lack of the connection between the firefighter's complaint of sex discrimination and retaliation and the underlying proceeding, Berkman against the City of New York, in which this application has been made.

Third, the defendant's [sic] contend that the Magistrate has erred in recommending the firefighter's continued detail to Engine Company 274 since, in

the defendants' view, that continued detail would be counterproductive because, according to the defendants, the necessary trust and confidence between the firefighter and other firefighters in the relatively small work force in which she operates does not now exist.

Finally, the defendants contend that the Magistrate erred as a matter of law in failing to balance the equities in determining whether to grant preliminary relief.

I will discuss each of these objections in the order in which I have just referred to them.

First, the Magistrate concluded that the Plaintiff would suffer irreparable harm since the firefighter's transfer from the company would, according to the Magistrate, have a chilling effect both on her and other members of the class in

this action in pursuing their efforts to secure non-discriminatory employment conditions.

The defendants argue that there is no evidence that removal of the firefighter from her detail, the Engine Company 274, for public safety reasons will have a chilling effect upon female firefighters who may be the victims of harassment.

I concluded that the defendant's [sic] argument is without merit.

On a review of the record I conclude that there is a substantial likelihood that during her detail to Engine Company 274 Firefighter Cannon was, in fact, subjected to sexual harassment and discriminatory treatment because of her position as a member of the class and her involvement in this litigation, and that her protest against the work

environment in which she was forced to operate was reason for the discrimination and harassment to which she was subjected.

The discrimination and sexual harassment that this record reflects are truly shocking.

The record of the hearing before Magistrate Caden amply documents a pattern of what can only be considered inhumane and inexcusable acts performed not only by other firefighters in the engine company, but also by the company's officers.

It is because of these circumstances and because of the repetition in this proceeding of what appeared in an earlier proceeding involving Firefighter Berkman and Gonzalez, that I have asked a representative of the United States Attorney's Office to be present. Because

of my concern that the defendants are unable to protect the civil rights of the members of the class. Accordingly, it is appropriate for the United States Attorney and the Federal Government to consider what steps by them may be necessary in order to assure the plaintiffs and other members of the class are not deprived of their civil rights by actions of the sort that are reflected in this record and the record of the earlier proceeding.

In that connection not only are proceedings for civil and criminal contempt appropriate for consideration but also enforcement of the Civil Rights Statutes which were originally enacted to permit steps to be taken against persons who deprive others of their civil rights by actions of the type that are reflected in this record.

Given the existence of sex discrimination and harassment of that sort in this record -- and I must note that no objection has been made to the Magistrate's findings and conclusions that such sex discrimination and harassment exists, given the existence of this record, it is difficult to understand an objection which says that a transfer of the only female member of this company, the one who is the victim of the harassment will not be regarded by other female members of the department as a continuation of harassment and have a decidedly chilling effect on others who seek to protect themselves from sexual harassment and discrimination of this sort.

The company is not a large one and why it is that the victim of the harassment is the one who is subjected to

administrative action and not the perpetrators escapes me.

The record reveals that there exists within the department ample means of communication between firefighters that will give a clear signal to other female firefighters, if this transfer is permitted to occur, that they must remain silent in the face of discrimination and harassment, or else, they too will be punished or administratively sanctioned, as Firefighter Cannon has apparently been punished or sanctioned.

As the Magistrate noted in his opinion, a chilling effect of this sort constitutes irreparable harm that money damages cannot adequately compensate.

I refer in that connection to Holt against Continental Group, Inc., decided by the Second Circuit in 1983 and

Sheehan against Purolator Courier decided by the same court in 1981.

The defendants argue that there would be no chilling effect as a result from Firefighter's [sic] Cannon's transfer since, "Uncontroverted evidence shows that the Fire Department would attempt to assign another female firefighter to Engine Company 274 in order to prevent any perception by male firefighters that misbehavior by male firefighters is an effective way to keep female firefighters out of the firehouse."

The contention, first of all, mischaracterizes the evidence. Chief O'Rourke, to whom reference is apparently made, testified to know [sic] more than, "We would consider putting another female firefighter in that company."

More importantly, however, it ignores the fact that in a situation in

which the victim is the one who is subjected to the administrative sanctions, it hardly suffices to avoid the chilling effect of those sanctions to put another victim in the place of the one who has just suffered.

In addition to the chilling effect that would result from failure to enjoin the proposed transfer, the Magistrate concluded that there would be irreparable injury in terms of loss of job skills and physical conditioning.

I do not accept that finding, since on the basis of the record established here, I am not persuaded that an off line assignment for fewer than nine months of this sort would result in a significant loss of firefighting skills or conditioning of the sort that could not be recovered.

I recognize that there are cases in which it is pointed out that degradation and humiliation in and of itself cannot constitute irreparable injury. I refer in that connection to Sampson against Morey and Stuart against The Immigration Naturalization Service.

However, the injury that is being spoken of here is not the simple degradation and humiliation of one woman subjected to this harassment but rather the effect that that degradation and humiliation will have on her in the future and on other similarly situated when, as now appears likely unless further steps are taken, continued discrimination and harassment occur.

In all events, this appears to me one of those extraordinary cases in which degradation and humiliation of one individual constitutes the kind of irrep-

arable injury for which the Court should take action.

I turn next to the jurisdictional point which I discuss in that order, because that is the order in which the objection was made. The claim here made is that there is a lack of sufficient nexus or connection between the discrimination to which Firefighter Cannon has been subjected and the underlying action seeking to invalidate a discriminatory physical examination by the Fire Department.

The only meaningful distinction that has been made between this case and the earlier case involving Firefighter Berkman and Firefighter Gonzalez, over which this Court exercised jurisdiction in 1983, is that in this case the victim of the harassment, Firefighter Cannon, did not testify as a witness during the

course of any of the underlying proceedings.

That difference, however, is one without material significance.

As the record here demonstrates Firefighter Cannon was one of the first women to take and pass the interim qualifying test. She received considerable notice in the press and among her fellow firefighters for her success in availing herself of the benefits that the litigation secured.

She was somewhat ironically afforded considerable attention at the time of her taking the oath of office, which was administered to her directly by the Mayor of this City.

I say "ironically" because it is, to me, a cruel irony that the attention paid by the defendant New York City to the firefighter at that time stands in

such marked contrast to the lack of attention paid to the conditions under which she has had to act within her firehouse once she was sworn in, went through her probationary period and began to serve the City as a firefighter.

Now, in the context of this litigation, I am not persuaded that it is necessary for one to be physically present in the courtroom or testify in litigation in order for a court to have jurisdiction to hear complaints of retaliation of this sort.

It is certainly conceivable that discrimination, sex discrimination, could occur within the Fire Department which would have no connection with this litigation.

However, it is apparent from the record that was made before the Magistrate in this case, that the discrimi-

nation to which Firefighter Cannon has been subjected is directly related to this litigation and to the firefighter's status and activities as a member of the class.

During the course of the hearings before the Magistrate, there was ample evidence that a number of the incidents in which the firefighter was involved arose out of discussions concerning this case.

In the fall of 1985 during a dinner at the firehouse, Firefighter Cannon, Lieutenant Brandenberger and Batallion [sic] Chief Griffen, among others, discussed an incident involving two other firefighters. When firefigher Cannon questioned how the firefighters' union could allow these firefighters to find themselves in the situation in which they did, Chief Griffen responded: "If

the union had any power, you wouldn't be sitting at this table having dinner with us right now."

When the Lieutenant and the other firefighters in the house laughed, Firefighter Cannon quite appropriately told Chief Griffin: "You're acting non-professional. You're putting me down in front of my officers, my co-workers."

The evidence before the Magistrate establishes at that point this superior officer said to the firefighter: "Shut up before you get in any more trouble."

Chief Griffin's offensive statements were not noted in the company journal.

At another dinner in January 1986, the Battalion Chief, according to the evidence before the Magistrate stated: "The women we have can't do the

job. They don't belong here. Any company that is riding with a woman is riding short. They should pay them to stay home. They should be hanging curtains."

This together with evidence before the Magistrate, that the incident which precipitated the April 8, 1986 dispute between Firefighter Cannon and Captain Drum, resulted from the Commissioner's order for a reevaluation of all women firefighters, apparently as a direct result of the demonstration by the firefighters union protesting this Court's decision relating to Exam 1162, persuades me that the retaliation and discrimination to which this firefighter has been subjected, arise directly out of this litigation, out of the firefighter's position as a class member and out of her efforts to enjoy the benefits conferred upon her by law in this case.

I turn next to the remedy. The defendants argue that even assuming a likelihood that the firefighter will succeed in the merits of her claim, the Magistrate erred in directing the firefighter's continued detail of [sic] Engine Company 274 since that remedy would be counterproductive.

The defendants base their argument on a series of cases in which the option of an award of money damages existed as an alternative to an order of reinstatement and in which an order directing the payment of back pay or money damages was chosen because of acrimony between the parties.

The application of those cases to this one could be debated.

In all events, the evidence here does not persuade me that the dissatisfaction and disharmony such as ex-

ists among a number of the firefighters in Engine Company 274 is of such a nature that it prevents the engine company from functioning when there are firefighting functions to be performed.

Almost all of the incidents of discrimination and harassment involved here involve the company in its off hours, that is, off firefighting hours, when it was not involved in actual firefighting. During actual firefighting there is no evidence here of inadequate performance, either by Firefighter Cannon or by anyone else.

I am persuaded that active and intelligent intervention by a commissioner and the administration of the City in this situation can preserve a functioning company without the chilling effect which would be involved in the

transferring of this firefighter from the company in which she has been assigned.

In all events, as I have noted earlier, there are certainly other avenues for relief which do not involve the sanctions being meted out on the victim of the disharmony.

Nothing I have said here should indicate that it is not appropriate for the parties to continue, if they have not already discussed manners and means by which the situation can be dealt with, with satisfaction to everyone involved [sic].

I have left a provision in the Order for the parties to consider whether there are other means of defusing this situation.

But it certainly does not seem to me the appropriate avenue of relief to have the one who is the victim of the

harassment be the one on whom all the burdens of solving it are placed.

That brings me to the final item, which is balance of hardships. Whether this was adequately examined by the Magistrate or not, my view of the record makes clear that the balance of hardships does not tip decidedly toward the defendants, and, since it is [sic] has also been mentioned, that the public consequences of an injunction in this case have not been demonstrated to be such as to argue against granting the injunction.

As I have already noted there is no showing made here that this company does not perform effectively when it is fighting fires.

On the other side of the balance there is the shocking image of a city agency and members of a city agency

unable to respect the directions of a court and unable to protect the civil rights of one of the employees of the agency. That sort of public disgrace, it seems to me, weighs in favor of the relief that is being sought here and outweighs any countervailing consideration with respect to the discomfort that will be created by people having to learn to live with each other and deal with each other as human beings.

Finally, the defendants have made an application to stay the issuance of the injunction pending appeal. From what I have said already, considering the factors that are appropriate for consideration with regard to that application, I am going to deny the stay pending appeal since I have little confidence that the moving party has likelihood of success on appeal, since the irreparable

injury that the City will suffer, if the stay is denied, appears to me slight and since the effect of issuing a stay and permitting a transfer on the public interest appears to me to be that the stay would hurt the public interest rather than further it.

Accordingly, I am going to sign an order of preliminary injunction and file it today.

You will note that the Order provides for service of the Order on the members of Engine Company 274, and failing proof of service with this Court it also gives notice for all people to whom the Order applies and to whose attention it is brought that disobedience to the Order may constitute civil or criminal contempt, and fine or imprisonment may be appropriate in the event civil or criminal contempt of court is committed.

THE COURT: Show the Order to the parties and file it.

(Proceedings concluded as above set forth.)

§ 2000e-2. Unlawful employment practices

(a) It shall be an unlawful employment practice for an employer --

(1) to fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment, because of such individual's race, color, religion, sex, or national origin; or

(2) to limit, segregate, or classify his employees or applicants for employment in any way which would deprive or tend to deprive any individual of employment opportunities or otherwise adversely affect his status as an employee, because of such individual's race, color, religion, sex, or national origin.

. . . .

(h) Notwithstanding any other provision of this subchapter, it shall not be an unlawful employment practice for an employer to apply different standards of compensation, or different terms, conditions, or privileges of employment pursuant to a bona fide seniority or merit system, or a system which measures earnings by quantity or quality of production or to employees who work in different locations, provided that such differences are not the result of an intention to discriminate because of race, color, religion, sex, or national origin, nor shall it be an unlawful employment practice for an employer to give and to act upon the results of any professionally developed ability test provided that such test, its administration or action upon the results is not de-

signed, intended or used to discriminate because of race, color, religion, sex or national origin. It shall not be an unlawful employment practice under this subchapter for any employer to differentiate upon the basis of sex in determining the amount of the wages or compensation paid or to be paid to employees of such employer if such differentiation is authorized by the provisions of section 206(d) of Title 29.

PART 1607--UNIFORM GUIDELINES ON EMPLOYEE SELECTION PROCEDURES (1978)

§ 1607.5 General standards for validity studies.

. . . .

B. Criterion-related, content, and construct validity. Evidence of the validity of a test or other selection procedure by a criterion-related validity study should consist of empirical data demonstrating that the selection procedure is predictive of or significantly correlated with important elements of job performance. See section 14B below. Evidence of the validity of a test or other selection procedure by a content validity study should consist of data showing that the content of the selection procedure is representative of important aspects of performance on the job for which the candidates are to be evaluated. See 14C below. Evidence of the validity

of a test or other selection procedure through a construct validity study should consist of data showing that the procedure measures the degree to which candidates have identifiable characteristics which have been determined to be important in successful performance in the job for which the candidates are to be evaluated. See section 14D below.

. . . .

G. Method of use of selection procedures. The evidence of both the validity and utility of a selection procedure should support the method the user chooses for operational use of the procedure, if that method of use has a greater adverse impact than another method of use. Evidence which may be sufficient to support the use of a selection procedure on a pass/fail (screening) basis may be insufficient to support the use of the

same procedure on a ranking basis under these guidelines. Thus, if a user decides to use a selection procedure on a ranking basis, and that method of use has a greater adverse impact than use on an appropriate pass/fail basis (see section 5H below), the user should have sufficient evidence of validity and utility to support the use on a ranking basis. See sections 3B, 14B(5) and (6), and 14C(8) and (9).

H. Cutoff scores. Where cutoff scores are used, they should normally be set so as to be reasonable and consistent with normal expectations of acceptable proficiency within the work force. Where applicants are ranked on the basis of properly validated selection procedures and those applicants scoring below a higher cutoff score than appropriate in light of such expectations have little or

no chance of being selected for employment, the higher cutoff score may be appropriate, but the degree of adverse impact should be considered.

§ 1607.14 Technical standards fcr validity studies.

. . . .

C. Technical standards for content validity studies--(1) Appropriateness of content validity studies. Users choosing to validate a selection procedure by a content validity strategy should determine whether it is appropriate to conduct such a study in the particular employment context. A selection procedure can be supported by a content validity strategy to the extent that it is a representative sample of the content of the job. Selection procedures which purport to measure knowledges, skills, or abilities may in certain circumstances be

justified by content validity, although they may not be representative samples, if the knowledge, skill, or ability measured by the selection procedure can be operationally defined as provided in section 14C(4) below, and if that knowledge, skill, or ability is a necessary prerequisite to successful job performance.

A selection procedure based upon inferences about mental processes cannot be supported solely or primarily on the basis of content validity. Thus, a content strategy is not appropriate for demonstrating the validity of selection procedures which purport to measure traits or constructs, such as intelligence, aptitude, personality, commonsense, judgment, leadership, and spatial ability. Content validity is also not an appropriate strategy when the selection

procedure involves knowledges, skills, or abilities which an employee will be expected to learn on the job.

(2) Job analysis for content validity. There should be a job analysis which includes an analysis of the important work behavior(s) required for successful performance and their relative importance and, if the behavior results in work product(s), an analysis of the work product(s). Any job analysis should focus on the work behavior(s) and the tasks associated with them. If work behavior(s) are not observable, the job analysis should identify and analyze those aspects of the behavior(s) that can be observed and the observed work products. The work behavior(s) selected for measurement should be critical work behavior(s) and/or important work behavior(s) constituting most of the job.

(3) Development of selection procedures. A selection procedure designed to measure the work behavior may be developed specifically from the job and job analysis in question, or may have been previously developed by the user, or by other users or by a test publisher.

(4) Standards for demonstrating content validity. To demonstrate the content validity of a selection procedure, a user should show that the behavior(s) demonstrated in the selection procedure are a representative sample of the behavior(s) of the job in question or that the selection procedure provides a representative sample of the work product of the job. In the case of a selection procedure measuring a knowledge, skill, or ability, the knowledge, skill, or ability being measured should be operationally defined. In the case of a se-

lection procedure measuring a knowledge, the knowledge being measured should be operationally defined as that body of learned information which is used in and is a necessary prerequisite for observable aspects of work behavior of the job. In the case of skills or abilities, the skill or ability being measured should be operationally defined in terms of observable aspects of work behavior of the job. For any selection procedure measuring a knowledge, skill, or ability the user should show that (a) the selection procedure measures and is a representative sample of that knowledge, skill, or ability; and (b) that knowledge, skill, or ability is used in and is a necessary prerequisite to performance of critical or important work behavior(s). In addition, to be content valid, a selection procedure measuring a skill or ability

should either closely approximate an observable work behavior, or its product should closely approximate an observable work product. If a test purports to sample a work behavior or to provide a sample of a work product, the manner and setting of the selection procedure and its level and complexity should closely approximate the work situation. The closer the content and the context of the selection procedure are to work samples or work behaviors, the stronger is the basis for showing content validity. As the content of the selection procedure less resembles a work behavior, or the setting and manner of the administration of the selection procedure less resemble the work situation, or the result less resembles a work product, the less likely the selection procedure is to be content

valid, and the greater the need for other evidence of validity.

(5) Reliability. The reliability of selection procedures justified on the basis of content validity should be a matter of concern to the user. Whenever it is feasible, appropriate statistical estimates should be made of the reliability of the selection procedure.

(6) Prior training or experience. A requirement for or evaluation of specific prior training or experience based on content validity, including a specification of level or amount of training or experience, should be justified on the basis of the relationship between the content of the training or experience and the conent of the job for which the training or experience is to be required or evaluated. The critical

consideration is the resemblance between the specific behaviors, products, knowledges, skills, or abilities in the experience or training and the specific behaviors, products, knowledges, skills, or abilities required on the job, whether or not there is close resemblance between the experience or training as a whole and the job as a whole.

(7) Content validity of training success. Where a measure of success in a training program is used as a selection procedure and the content of a training program is justified on the basis of content validity, the use should be justified on the relationship between the content of the training program and the content of the job.

(8) Operational use. A selection procedure which is supported on the basis of content validity may be used for

a job if it represents a critical work behavior (i.e., a behavior which is necessary for performance of the job) or work behaviors which constitute most of the important parts of the job.

(9) Ranking based on content validity studies. If a user can show, by a job analysis or otherise, that a higher score on a content valid selection procedure is likely to result in better job performance, the results may be used to rank persons who score above minimum levels. Where a selection procedure supported solely or primarily by content validity is used to rank job candidates, the selection procedure should measure those aspects of performance which differentiate among levels of job performance.